HUMAN BIOLOGY
a record of research

MAY, 1950

VOL. 22 No. 2

# STATISTICAL MODELS BEARING ON THE SEMANTICS OF CORRELATION

## III. EXPECTATION-ACCOMPLISHMENT MODELS

BY LANCELOT HOGBEN, F.R.S. AND KENNETH W. KEMP
*Department of Medical Statistics, University of Birmingham*

### 1. INTRODUCTION

THE present series of communications has as its main objective to clarify:

(a) in what sense and in what terms we may partition the variance of the distribution of a system of scores into components referable (*explained*) or otherwise (*unexplained*) to a source of variation inherent in or contributory to the distribution of a second system of scores correlated therewith;

(b) in what circumstances linear regression may arise and how far it is indicative of an intrinsic linear law of concomitant variation descriptive of such a dual system of scoring.

The two previous contributions have shewn that it is necessary to distinguish between three kinds of logical relationship respectively distinguished as *consequence, concurrence* and *constraint*. It is not possible to make explicit in these terms the nature of the relationship from which correlation between two systems of scoring arises by recourse to the

customary geometrical approach invoking the method of least squares or the properties of a bivariate normal distribution; but it is possible to do so by exact algebraic methods which we can visualise *vis à vis* appropriate models. In the notation of this procedure certain relations between components of variance and relations definitive of linear regression become manifest as tautologies of any score-frequency grid regardless of any causal meaning we may attach to them by invoking the theory of probability. It is therefore not surprising if we later discover that they admit of no unique interpretation, when the end in view is to disclose the mechanism inherent in a system of correlated variates.

Nor is it wholly unnecessary to remind ourselves that this is so. For it is safe to say that a relatively small proportion of biologists or sociologists who habitually invoke procedures generically referred to as analysis of variance and analysis of covariance are cognisant of the logical postulates implicit in their mathematical rationale. Indeed one must search far afield for an unequivocal statement of two general conclusions set out in § 3 of Part I of this series, *viz.*:

(i) the familiar additive relationship between total variance, inter-class (class mean) variance and intra-class (mean class) variance in a two-dimensional lay-out of scores and frequencies is inherent in the lay-out of the numbers themselves regardless of what sources of variation contribute thereto;

(ii) the particular significance the *within*-class and *between*-class components of the total variance of one or other set of scores in such a bivariate system enjoy as definitive of linear regression in one or other dimension of the grid is likewise deducible from its structure without recourse to any specifically statistical assumptions.

If it is still necessary to state such conclusions with emphasis, it would be less than fair to make no acknowledgement to Churchill Eisenhart[1] (1947), who evidently also recognizes the need to do so in the following remarks. For the *italics* are his own:

> ***I repeat: All of the familiar formulas and procedures for evaluating component " sums of squares " that add up to the " total sum of squares " are based on algebraic identities, and are valid as descriptions of properties of the data whatever the interpretation of the numbers involved.*** . . . Similarly, all of the familiar formulas and procedures for evaluating regression coefficients and the sum of squared deviations from the fitted

[1] Churchill Eisenhart (1947), *Biometrics* (Amer. Stat. Assn.), vol. 3, pp. 7-8.

**regression, when the fitting is by the method of least squares, are based upon algebra and calculus, and the results obtained are valid as descriptions of properties of the data in hand, whatever the interpretation of the numbers involved. *In summary, when the formulas and procedures of analysis of variance are used merely to summarize properties of the data in hand, no assumptions are needed to validate them. On the other hand, when analysis of variance is used as a method of statistical inference, for inferring properties of the "population" from which the data in hand were drawn, then certain assumptions, about the "population" and the sampling procedure by means of which the data were obtained, must be fulfilled if the inferences are to be valid.***

In this communication it is our concern to exhibit the bearing of certain sampling procedures on a meaningful partition of variance and cognate issues by recourse to a class of models whose common feature is that the column border A-score ($x_a$) is a linear function of the expectation of the player's actual score ($x_b$) in a single trial. We shall consider three variants on this theme:

A. The player draws the same number ($s$) of balls from each of $u$ urns, each containing $n$ balls of which $x_a$ are red, the player's score ($x_b$) at a given trial being the number of red balls in the $s$-fold sample;

B. The player draws from one and the same urn of $n$ balls, of which $pn$ are red, samples of $x_a$ balls recording his score $x_b$ in a sample of $x_a$ items as the number of red balls therein;

C. The player chooses one card from each of a set of packs containing different numbers ($x_a$) of cards *consecutively* numbered $1, 2, 3, \cdots$ etc., and records as his score ($x_b$) the number on the card chosen.

In the terminology of the previous contributions of this series the method of scoring the player's luck in A or B alike is *taxonomic*, that of C is *representative*. Each exemplifies an essentially *consequential* relationship. The first two alone admit of a meaningful partition of variance with respect to the source of variation; and the partition appropriate to each model is distinctive. It is the common property of all three that they exhibit linear regression in one dimension only, except when we impose the special condition dealt with in § 6 below, and illustrated in Example 1*a* of § 2.

When speaking of a *causal* partition of variance, it is desirable to distinguish between a balance sheet whose items are unbiassed estimates based on sample data and one which exhibits overall relations between

components of variance in the putative universe. The possibility of defining a balance sheet of variance in the latter sense is the implicit rationale of a meaningful partition of variance based on the information

**SCORE - GRID**
**for 2 criteria of classification**

| | A | B | C | $M_r$ | $V_r$ |
|---|---|---|---|---|---|
| P | 3 | 1 | 5 | 3 | 8/3 |
| Q | 1 | 2 | 3 | 2 | 2/3 |
| R | 4 | 3 | 5 | 4 | 2/3 |
| $M_c$ | 8/3 | 2 | 13/3 | | |
| $V_c$ | 14/9 | 2/3 | 8/9 | | |

$M(V_r) + V(M_r) = V = M(V_c) + V(M_c)$

$4/3 + 2/3 = 2 = 26/27 + 28/27$

**FREQUENCY GRID**
**for 1 criterion**
**(A-C dimensional)**

| | A | B | C |
|---|---|---|---|
| 1 | 1/9 | 1/9 | 0 |
| 2 | 0 | 1/9 | 0 |
| 3 | 1/9 | 1/9 | 1/9 |
| 4 | 1/9 | 0 | 0 |
| 5 | 0 | 0 | 2/9 |
| $M_{ba}$ | 8/3 | 2 | 13/3 |
| $V_{ba}$ | 14/9 | 2/3 | 8/9 |

$V(M_{ba}) + M(V_{ba}) = V$

$26/27 + 28/27 = 2$

**FREQUENCY GRID**
**for 1 criterion (P-R dimensional)**

| | 1 | 2 | 3 | 4 | 5 | $M_{ab}$ | $V_{ab}$ |
|---|---|---|---|---|---|---|---|
| P | 1/9 | 0 | 1/9 | 0 | 1/9 | 3 | 8/3 |
| Q | 1/9 | 1/9 | 1/9 | 0 | 0 | 2 | 2/3 |
| R | 0 | 0 | 1/9 | 1/9 | 1/9 | 4 | 2/3 |

$V(M_{ab}) + M(V_{ab}) = V$

$2/3 + 4/3 = 2$

FIG. 1. SCORE GRID AND FREQUENCY GRID

a sample supplies. In this one, as in previous contributions of the series, our only concern is with partition of variance *in the universe as a whole.*

This distinction directs attention to a difference between two ways of setting out data gridwise. The explicit cell entries of the correlation

grid of our models here and elsewhere exhibit the theoretical frequencies of scores or functions thereof laid out at the head of the column and margin of the rows. What we speak of as a Fisher score-grid of $rc$ cells exhibits as each cell entry a score of frequency $1/rc$. With respect to one criterion of classification (Fig. 1) we may lay out such a set of data as a grid whose cell entries are theoretical frequencies referable to one set of border-scores, so that the fundamental tautology of a correlation grid is valid for a score grid, *i. e.* $M(V_r) + V(M_r) = V$ if we label by border scores the rows, or alternatively $M(V_c) + V(M_c) = V$

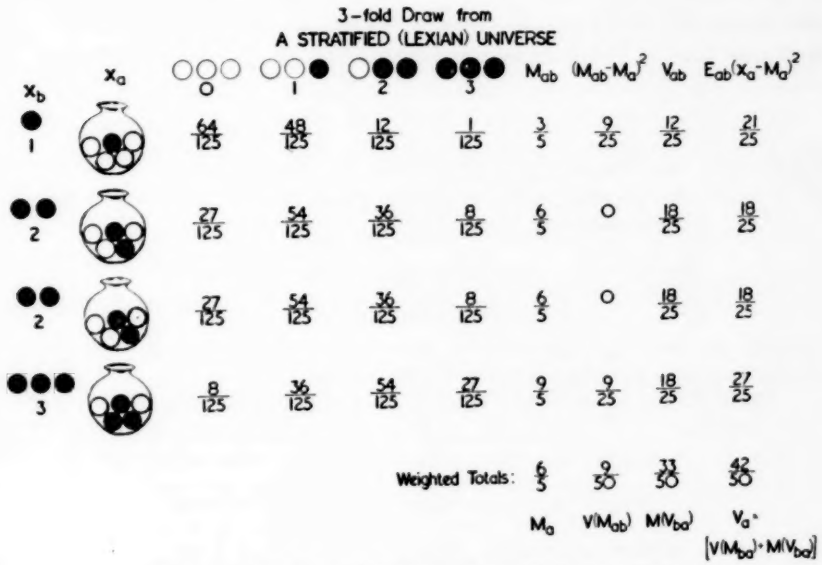


FIG. 2. THE CLASS A MODEL (WITH REPLACEMENT)

if we label by border scores the columns. For one criterion of classification, we may therefore use the score-grid to exhibit the score distribution of a universe; but we can represent a universe stratified in virtue of two criteria of classification only by a 3-dimensional grid. A score-grid lay-out which does in fact involve two criteria of classification is *ipso facto* a description of a single sample only.

## 2. NUMERICAL EXAMPLES OF EXPECTATION-ACCOMPLISHMENT MODELS

Models of Class A which we designate below as *Lexian* models include two types according as we do (Fig. 2) or do not impose the condition

that the player replaces each ball chosen before drawing another. When there is replacement the sample distribution is given by the terms of the binomial

$$(\overline{n - x_a} + x_a)^r / n^r.$$

When there is no replacement the definitive binomial is, of course:

$$(\overline{n - x_a} + x_a)^{(r)} / n^{(r)}.$$

A special case, discussed in § 6 below, arises when there is no replacement, if $u = (n + 1)$, and $x_a = 0, 1, 2 \cdots n$, as illustrated by Example 1*a*.

*Example* 1*a*. Six urns each contain 5 balls of which $0, 1, 2, \ldots, 5$ are respectively red. The player may choose *without replacement* 3 balls from each urn.

| | | RED BALLS PER URN ($x_a$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total | Mean | Variance |
| RED BALLS PER SAMPLE ($x_b$) | 0 | 10 | 4 | 1 | 0 | 0 | 0 | 15 | 2/5 | 28/75 |
| | 1 | 0 | 6 | 6 | 3 | 0 | 0 | 15 | 9/5 | 42/75 |
| | 2 | 0 | 0 | 3 | 6 | 6 | 0 | 15 | 16/5 | 42/75 |
| | 3 | 0 | 0 | 0 | 1 | 4 | 10 | 15 | 23/5 | 28/75 |
| Total | | 10 | 10 | 10 | 10 | 10 | 10 | 60 | 5/2 | 35/12 |
| Mean | | 0 | 3/5 | 6/5 | 9/5 | 12/5 | 3 | 3/2 | — | — |
| Variance | | 0 | 12/5 | 18/5 | 18/5 | 12/5 | 0 | 5/4 | — | — |

$$M_a = 5/2 \quad ; \quad k_{ba} = 3/5 \quad ; \quad M_b = 3/2$$

$$V(M_{ab}) = 49/20 \quad ; \quad M(V_{ab}) = 28/60 \quad ; \quad V(M_{ab}) + M(V_{ab}) = 35/12 = V_a$$

$$V(M_{ba}) = 21/20 \quad ; \quad M(V_{ba}) = 1/5 \quad ; \quad V(M_{ba}) + M(V_{ba}) = 5/4 = V_b$$

$$Cov\,(x_a, x_b) = 7/4 = k_{ab} \cdot V_a \quad ; \quad V(M_{ba})/V_b = 21/25 = r_{ab}^2$$

*Example* 1*b*. The constitution of the urns and the size of the sample is as for Example 1*a*, but the player *replaces* each ball before drawing another from the same urn.

RED BALLS PER URN ($x_a$)

| | | 0 | 1 | 2 | 3 | 4 | 5 | Total | Mean | Variance |
|---|---|---|---|---|---|---|---|---|---|---|
| RED BALLS PER SAMPLE ($x_b$) | 0 | 125 | 64 | 27 | 8 | 1 | 0 | 225 | 146/225 | $37184/(225)^2$ |
| | 1 | 0 | 48 | 54 | 36 | 12 | 0 | 150 | 312/150 | $19656/(150)^2$ |
| | 2 | 0 | 12 | 36 | 54 | 48 | 0 | 150 | 438/150 | $19656/(150)^2$ |
| | 3 | 0 | 1 | 8 | 27 | 64 | 125 | 225 | 979/225 | $37184/(225)^2$ |
| Total | | 125 | 125 | 125 | 125 | 125 | 125 | 750 | 5/2 | 35/12 |
| Mean | | 0 | 3/5 | 6/5 | 9/5 | 12/5 | 3 | 3/2 | — | — |
| Variance | | 0 | 12/25 | 18/25 | 18/25 | 12/25 | 0 | 29/20 | — | — |

$M_a = 5/2$ ; $k_{ba} = 3/5$ ; $M_b = 3/2$

$V(M_{ab}) = 717703/(750)(450)$ ; $M(V_{ab}) = 66668/(375)(225)$ ; $V_a = 35/12$

$V(M_{ba}) = 21/20$ ; $M(V_{ba}) = 2/5$ ; $V_b = 29/20$

$$Cov\,(x_a, x_b) = 7/4 = k_{ba} \,.\, V_a$$

$$V(M_{ba})/V_b = r_{ab}^2 = 21/29 = \frac{s(u+1)}{(s-1)(u+1) + 3(2n-u+1)}$$

*Example* 1*c*. The player chooses as before 3 balls from the same six different sorts of urns each containing 5 balls, replacing each ball drawn before taking another. There are however 32 urns instead of 6, the $x_a$ score distribution being binomial in accordance with the terms of $(\frac{1}{2} + \frac{1}{2})^5$, *i. e.*, the urns are as follows:

| No. of red balls per urn | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| No. of urns | 1 | 5 | 10 | 10 | 5 | 1 |

RED BALLS PER URN ($x_a$)

| | | 0 | 1 | 2 | 3 | 4 | 5 | Total | $M_{ab}$ | $V_{ab}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| RED BALLS PER SAMPLE ($x_b$) | 0 | 25 | 64 | 54 | 16 | 1 | 0 | 160 | 7/5 | 79/100 |
| | 1 | 0 | 48 | 108 | 72 | 12 | 0 | 240 | 11/5 | 66/100 |
| | 2 | 0 | 12 | 72 | 108 | 48 | 0 | 240 | 14/5 | 66/100 |
| | 3 | 0 | 1 | 16 | 54 | 64 | 25 | 160 | 18/5 | 79/100 |
| Total | | 25 | 125 | 250 | 250 | 125 | 25 | 800 | — | — |
| $M_{ba}$ | | 0/5 | 3/5 | 6/5 | 9/5 | 12/15 | 15/5 | — | — | — |
| $V_{ba}$ | | 0 | 12/25 | 18/25 | 18/25 | 12/25 | 0 | — | — | — |

$V(M_{ab}) = 269/500$ ; $M(V_{ab}) = 89/125$ ; $V(M_{ab}) + M(V_{ab}) = 5/4 = V_a$

$V(M_{ba}) = 9/20$ ; $M(V_{ba}) = 3/5$ ; $V(M_{ba}) + M(V_{ba}) = 21/20 = V_b$

$$V(M_{ba})/V_b = r_{ab}^2 = 3/7 = \frac{s \,.\, V_a}{(s-1)V_a + M_a(n - M_a)}$$

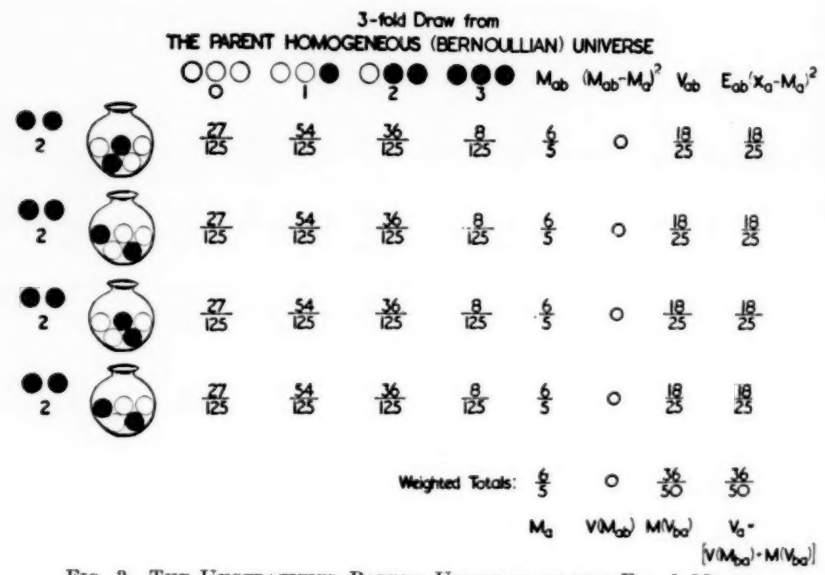


FIG. 3. THE UNSTRATIFIED PARENT UNIVERSE OF THE FIG. 1 MODEL

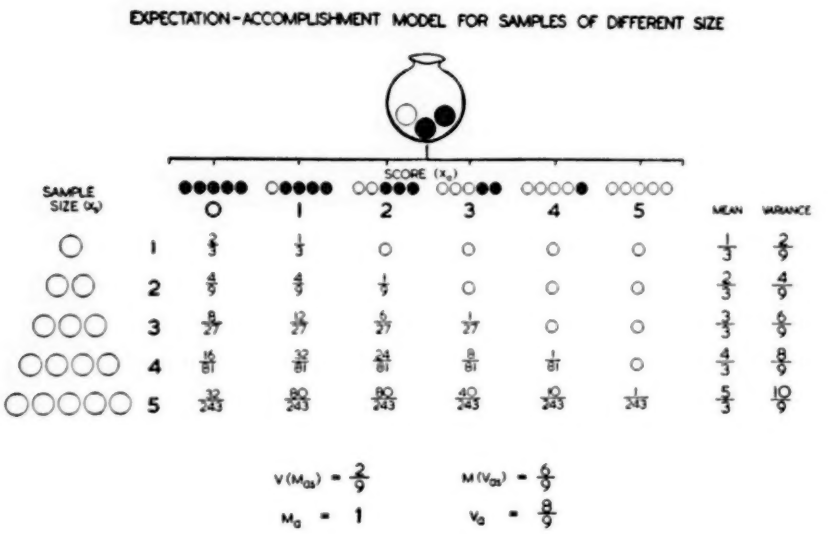


FIG. 4. THE CLASS B MODEL

Models of Class B (Fig. 4) like those of Class A likewise accommodate two types with respect to the condition of replacement.

*Example 2a.* The player draws from an urn containing 12 balls of which 8 are red, and separately records the result of taking samples of 1, 2, 3, 4 or 5 balls *without* replacement.

SAMPLE SIZE ($x_a$)

| PLAYER'S SCORE ($x_b$) | 1 | 2 | 3 | 4 | 5 | Total | $M_{ab}$ | $V_{ab}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 165 | 45 | 9 | 1 | 0 | 220 | 13/10 | 351/1100 |
| 1 | 330 | 240 | 108 | 32 | 5 | 715 | 18/10 | 5824/7150 |
| 2 | 0 | 210 | 252 | 168 | 70 | 700 | 157/50 | 2301/2500 |
| 3 | 0 | 0 | 126 | 224 | 210 | 560 | 83/20 | 231/400 |
| 4 | 0 | 0 | 0 | 70 | 175 | 245 | 33/7 | 10/49 |
| 5 | 0 | 0 | 0 | 0 | 35 | 35 | 5 | 0 |
| Total | 495 | 495 | 495 | 495 | 495 | 2475 | 3 | 2 |
| $M_{ba}$ | 2/3 | 4/3 | 6/3 | 8/3 | 10/3 | 2 | — | — |
| $V_{ba}$ | 22/99 | 40/99 | 54/99 | 64/99 | 70/99 | 138/99 | — | — |

$$M_a = 3 \quad ; \quad k_{ba} = 2/3 \quad ; \quad M_b = 2$$

$$M(V_{ab}) = 41757/61875 \quad ; \quad M(V_{ba}) = 50/99 \quad ; \quad V(M_{ab}) = 81993/61875 \quad ;$$

$$V(M_{ba}) = 8/9 = k_{ba}^2 \,.\, V_a$$

$$M(V_{ab}) + V(M_{ab}) = 2 = V_a \quad ; \quad M(V_{ba}) + V(M_{ba}) = 138/99 = V_b$$

$$V(M_{ab})/V_a = 27331/41250 \quad ; \quad V(M_{ba})/V_b = 44/69 = r_{ab}^2$$

*Example 2b.* The player replaces each ball before drawing another but the set-up is otherwise as for Example 2*a*, so that $k_{ba} = 2/3$.

SAMPLE SIZE ($x_a$)

| PLAYER'S SCORE ($x_b$) | 1 | 2 | 3 | 4 | 5 | Total | $M_{ab}$ | $V_{ab}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 81 | 27 | 9 | 3 | 1 | 121 | 179/121 | $9462/(121)^2$ |
| 1 | 162 | 108 | 54 | 24 | 10 | 358 | 351/179 | $35754/(179)^2$ |
| 2 | 0 | 108 | 108 | 72 | 40 | 328 | 257/82 | $6849/(82)^2$ |
| 3 | 0 | 0 | 72 | 96 | 80 | 248 | 125/31 | $588/(31)^2$ |
| 4 | 0 | 0 | 0 | 48 | 80 | 128 | 37/8 | $15/(8)^2$ |
| 5 | 0 | 0 | 0 | 0 | 32 | 32 | 5 | 0 |
| Total | 243 | 243 | 243 | 243 | 243 | 1215 | 3 | 2 |
| $M_{ba}$ | 2/3 | 4/3 | 6/3 | 8/3 | 10/3 | 2 | — | — |
| $V_{ba}$ | 2/9 | 4/9 | 6/9 | 8/9 | 10/9 | 14/9 | — | — |

$$M(V_{ab}) = 0.767 \quad ; \quad M(V_{ba}) = 6/9 \quad ; \quad V(M_{ab}) = 1.233 \quad ;$$

$$V(M_{ba}) = 8/9 = k_{ba}^2 \,.\, V_a$$

$$M(V_{ab}) + V(M_{ab}) = 2 = V_a \quad ; \quad M(V_{ba}) + V(M_{ba}) = 14/9 = V_b$$

$$V(M_{ab})/V_a = 0.617 \,; \qquad V(M_{ba})/V_b = 4/7 = r_{ab}^2$$

In each numerical illustration of Class C, designated below as the *unit sample rectangular model* (Fig. 5), the player will take one card

THE RECTANGULAR UNIT SAMPLE MODEL

| Size of Pack ($x_a$) | 1 | 2 | 3 | 4 | 5 | 6 | Mean ($M_{xa}$) | Variance ($V_{xa}$) |
|---|---|---|---|---|---|---|---|---|
| Expected Score ($x_b$) of Unit Sample | | | | | | | | |
| 1 | 1 | $\frac{1}{2}$ | $\frac{1}{3}$ | $\frac{1}{4}$ | $\frac{1}{5}$ | $\frac{1}{6}$ | 2·449 | 2·574 |
| 2 | 0 | $\frac{1}{2}$ | $\frac{1}{3}$ | $\frac{1}{4}$ | $\frac{1}{5}$ | $\frac{1}{6}$ | 3·448 | 1·902 |
| 3 | 0 | 0 | $\frac{1}{3}$ | $\frac{1}{4}$ | $\frac{1}{5}$ | $\frac{1}{6}$ | 4·210 | 1·219 |
| 4 | 0 | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{5}$ | $\frac{1}{6}$ | 4·865 | ·657 |
| 5 | 0 | 0 | 0 | 0 | $\frac{1}{5}$ | $\frac{1}{6}$ | 5·455 | ·248 |
| 6 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ | 6 | 0 |
| Mean ($M_{xb}$) | 1 | 1·5 | 2 | 2·5 | 3 | 3·5 | | |
| Variance ($V_{xb}$) | 0 | $\frac{3}{12}$ | $\frac{8}{12}$ | $\frac{15}{12}$ | $\frac{24}{12}$ | $\frac{35}{12}$ | | |

FIG. 5. THE CLASS C MODEL

from a set of card packs containing 1, 2, 3, up to 6 cards of the club suit starting with the ace. Thus the smallest pack consists of an ace only, the 2-fold pack of the ace and 2, the largest pack containing the ace, 2, 3, 4, 5 and 6 of clubs. The score $x_a$ is the number of cards per pack, the player's score being $x_b$ clubs in the unit trial from any particular pack. The maximum score ($N$) of the player is therefore 6.

*Example 3a.* There are six packs in all, so that the player takes one card from each pack in a single set of trials. Thus the A-score distribution is rectangular like that of $x_b$ within a column.

NO. OF CARDS PER PACK ($x_a$)

| PLAYER'S SCORE ($x_b$) | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| 1 | 120 | 60 | 40 | 30 | 24 | 20 | 294 |
| 2 | — | 60 | 40 | 30 | 24 | 20 | 174 |
| 3 | — | — | 40 | 30 | 24 | 20 | 114 |
| 4 | — | — | — | 30 | 24 | 20 | 74 |
| 5 | — | — | — | — | 24 | 20 | 44 |
| 6 | — | — | — | — | — | 20 | 20 |
| Total | 120 | 120 | 120 | 120 | 120 | 120 | 720 |
| Mean | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 2.25 |
| Variance | 0 | 1/4 | 2/3 | 5/4 | 2.0 | 35/12 | 3.5 |

$$M_a = 3.5 \quad ; \quad k_{ba} = 1/2 \quad ; \quad M_b = 2.25 \quad ; \quad V_a = 35/12$$

$$V(M_{ba}) = 35/48 \quad ; \quad M(V_{ba}) = 85/72 \quad ; \quad V_b = 275/144 = V(M_{ba}) + M(V_{ba})$$

$$k_{ba}^2 \cdot V_a = 35/48 = V(M_{ba}) \quad ; \quad k_{ba} \cdot V_a = 35/24 = Cov\,(x_a, x_b)$$

$$V(M_{ba})/V_b = 21/55 = r_{ab}^2 = \frac{3(N+1)}{(7N+13)}$$

*Example 3b.* A set of unit trials involves choice of one card from each of 243 packs, the A-score distribution being defined by successive terms of the binomial $(2/3 + 1/3)^5$, *i. e.*, the set-up is as follows:

| No. of cards per pack | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| No. of packs | 32 | 80 | 80 | 40 | 10 | 1 |

As before therefore $N = 6$ and in (5.5) below $q = 2/3 = (1 - p)$.

NO. OF CARDS PER PACK ($x_a$)

| PLAYER'S SCORE ($x_b$) | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| 1 | 192 | 240 | 160 | 60 | 12 | 1 | 665 |
| 2 | — | 240 | 160 | 60 | 12 | 1 | 473 |
| 3 | — | — | 160 | 60 | 12 | 1 | 233 |
| 4 | — | — | — | 60 | 12 | 1 | 73 |
| 5 | — | — | — | — | 12 | 1 | 13 |
| 6 | — | — | — | — | — | 1 | 1 |
| Total | 192 | 480 | 480 | 240 | 60 | 6 | 1458 |
| Mean | 1 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 11/6 |

$$k_{ba} = 1/2 \quad ; \quad V_a = 10/9 = (s-1)pq$$

$$V(M_{ba}) = 5/18 = k_{ba}^2 \cdot V_a \quad ; \quad V_b = 95/108$$

$$V(M_{ba})/V_b = 6/19 = r_{ab}^2 = \frac{3q}{6 + (N-5)p}$$

### 3. LEXIAN MODELS

We here designate as *Lexian*, models of Class A in § 1 regardless of whether we do or do not impose the replacement condition. We suppose that there are $u$ urns containing an equal number ($n$) of balls, some red, some black, the proportion of red balls in the $a$-th urn being $p_a$ and the actual number ($np_a$) being the A (column border) score ($x_a$) of the grid. The player draws a sample of $s$ balls from each urn (with or without replacement) and records as the B (row border) score ($x_b$) the number of red balls in each sample. The symbol $M_a$ stands for the mean number of red balls per urn. With or without replacement, the mean score of an $s$-fold sample from the $a$-th urn is $sp_a$, so that:

$$p_a = \frac{x_a}{n} \quad ; \quad M_{ba} = sp_a = \frac{s}{n} x_a \quad ; \quad M_b = \frac{s}{n} E(x_a) = \frac{s}{n} M_a \quad (3.1)$$

$$\therefore M_{ba} - M_b = \frac{s}{n} (x_a - M_a)$$

$$\therefore k_{ba} = \frac{s}{n} .$$

Thus regression of the player's score on the A-score is necessarily linear. It is then a property of the score-frequency grid, as shown in (2.18-2.20) of § 2 in Part I that

$$V(M_{ba}) = V_b - M(V_{ba}) = k_{ba}^2 \cdot V_a = r_{ab}^2 V_b .$$

It will suffice to develop $r_{ab}^2$ in accordance with the condition that there is no replacement, since the alternative case is deducible therefrom as a limiting condition. If the player draws $s$ balls simultaneously from an urn containing $x_a$ red balls and $b_a = (n - x_a)$ black ones, the frequency function for the $x_b$ score is $(x_a + b_a)^{(s)}/n^{(s)}$; and its variance with respect to such an urn is given by

$$V_{ba} = \frac{s(n-s)}{n-1} p_a(1-p_a) = \frac{s(n-s)}{n^2(n-1)} x_a(n-x_a).$$

Now $M(V_{ba}) = E_a(V_{ba})$ and

$$\begin{aligned} E_a[x_a(n-x_a)] &= nM_a - E_a(x_a^2) \\ &= nM_a - (V_a + M_a^2) \\ &= M_a(n - M_a) - V_a, \end{aligned}$$

$$\therefore M(V_{ba}) = \frac{s(n-s)}{n^2(n-1)} [M_a(n - M_a) - V_a] \qquad (3.2)$$

$$\therefore V_b = \frac{s^2}{n^2} V_a + \frac{s(n-s)}{n(n-1)} M_a - \frac{s(n-s)}{n^2(n-1)} V_a - \frac{s(n-s)}{n^2(n-1)} M_a^2.$$

Whence we have:

$$r_{ab}^2 = \frac{s^2}{n^2}\frac{V_a}{V_b} = \frac{V_a \cdot s^2/n^2}{V_a \cdot s^2/n^2 + M(V_{ba})} .$$

In virtue of (3.2), the denominator of this expression is:

$$\frac{s(s-1)}{n(n-1)} V_a + \frac{s(n-s)}{n^2(n-1)} M_a(n - M_a),$$

whence

$$r_{ab}^2 = \frac{s(n-1)V_a}{n(s-1)V_a + (n-s)M_a(n-M_a)} . \qquad (3.3)$$

When $n = s$ in (3.3), $(n-s) = 0$ and $s(n-1) = n(s-1)$, so that $r_{ab}^2 = 1$. This is necessarily so, since the player then empties each urn. If $n$ is very large in comparison with $s$, we may write $(n-1) \simeq n \simeq (n-s)$, so that (3.3) reduces to:

$$r_{ab}^2 = \frac{s \cdot V_a}{(s-1)V_a + M_a(n - M_a)} . \qquad (3.4)$$

From a formal viewpoint it is usually immaterial whether we impose the restriction of replacement regardless of the size of the universe ($n$) or consider that $n$ is indefinitely large compared with $s$. If so, (3.4) is definitive of the *replacement* condition; but a peculiarity of this set-up is that either (3.3) or (3.4) reduce to zero as $n$ approaches infinity in virtue of the factor $(n - M_a)$ in the denominator. For that reason, it is appropriate to develop (3.4) independently:

$$V_{ba} = sp_a(1-p_a) = \frac{s}{n^2} \cdot x_a(n - x_a) = \frac{sx_a}{n} - \frac{sx_a^2}{n^2};$$

$$\therefore M(V_{ba}) = \frac{s}{n} E_a(x_a) - \frac{s}{n^2} E_a(x_a^2)$$

$$= \frac{s}{n} M_a - \frac{s}{n^2}(V_a + M_a^2);$$

$$\therefore V_b = \frac{s^2}{n^2} V_a + \frac{s}{n} M_a - \frac{s}{n^2}(V_a + M_a^2)$$

$$= \frac{s(s-1)}{n^2} V_a + \frac{s}{n^2} M_a(n - M_a),$$

whence we obtain (3.4). Both (3.3) and (3.4) uniquely depend on the distribution of the $\Lambda$-score, *i. e.*, that of the urn composition. If

the $u$ urns are all different and the scores run *consecutively* by unit steps as in Examples 1*a* and 1*b*, the distribution is rectangular, and

$$V_a = (u^2 - 1)/12.$$

If the minimum number of red balls in an urn is $m$, the mean value $(M_a)$ of $x_a$ is then $m + \frac{1}{2}(u - 1)$; and if $m = 0$, (3.4) becomes

$$r_{ab}{}^2 = \frac{s(u+1)}{(s-1)(u+1) + 3(2n - u + 1)}.$$

We may now proceed to enquire in what sense we can partition $V_b$, the variance of the player's score distribution, into two components $V_E$ and $V_U$ respectively definitive of *explained* and *unexplained* variation, *i. e.* of variation arising from and independent of the source, *viz.*, the circumstance that the urns do not all contain the same number of balls. In the first place, we note that we can give the issue so defined a unique meaning only if we regard the individual urns as sub-samples of a single universe which we can reconstruct on a Bernoullian basis, *i. e.* as a homogeneous system, by mixing the contents of the $u$ urns. How much of the variance of the B-score distribution then arises from the single relevant circumstance that the universe is *stratified* in virtue of the heterogeneity of the individual urns, we may then assign by deducting what the variance of the player's score would be if allowed to take $s$-fold samples from the reconstituted (Bernoullian) universe of $un$ balls. Though there is nothing arbitrary about such a definition of explained variation, it is important at the outset to recognise that it admits a unique solution only in virtue of an arbitrary property of the model, *viz.*, that each sub-universe contains the same number $(n)$ of items. With due regard to this limitation we proceed as follows.

To assign a value to the unexplained component $(V_u)$ of variance, we note that the homogeneous universe consists of $un$ balls of which $\Sigma x_a$ are red. If we denote the 1-fold trial expectation of success by $p_u$, we therefore have

$$p_u = \frac{\Sigma x_a}{un} = \frac{1}{n}\Sigma\frac{x_a}{u} = \frac{M_a}{n}.$$

When there is *no* replacement:

$$V_U = \frac{s(un-s)}{un-1} p_u q_u$$

$$= \frac{s(un-s)}{(un-1)} \cdot \frac{M_a}{n}\left(1 - \frac{M_a}{n}\right). \qquad (3.5)$$

In sampling from the stratified universe as above we found that

$$M(V_{ba}) = \frac{s(n-s)}{n-1} \cdot \frac{M_a}{n}\left(1-\frac{M_a}{n}\right) - \frac{(n-s)}{s(n-1)}\ V(M_{ba}),$$

$$\therefore \frac{M_a}{n}\left(1-\frac{M_a}{n}\right) = \frac{n-1}{s(n-s)} M(V_{ba}) + \frac{1}{s^2} V(M_{ba}).$$

Whence from (3. 5)

$$V_U = \frac{(un-s)(n-1)}{(un-1)(n-s)} M(V_{ba}) + \frac{(un-s)}{s(un-1)} V(M_{ba}). \qquad (3.6)$$

By definition $V_b = V_U + V_E$ and hence

$$V_U + V_E = M(V_{ba}) + V(M_{ba}),$$

$$\therefore V_E = \left[1 - \frac{(un-s)(n-1)}{(un-1)(n-s)}\right] M(V_{ba}) + \left[1 - \frac{(un-s)}{s(un-1)}\right] V(M_{ba}). \qquad (3.7)$$

If we put $(un-1) \simeq (un-s)$ and $(n-1) \simeq (n-s)$, we obtain the appropriate expressions when the condition of replacement holds good, and

$$V_U = M(V_{ba}) + \frac{1}{s} V(M_{ba}), \qquad (3.8)$$

$$V_E = \frac{s-1}{s} V(M_{ba}). \qquad (3.9)$$

In seeking for a partition of variance referable to the source, *viz.* stratification of the universe, we have here explored the result of removing the source of variation by assigning the variance of the player's score in sampling from a parent (Fig. 3) Bernoullian universe, *i. e.* by pooling the contents of the urns and extracting $s$-fold samples from the composite urn so constituted. Provided $x_a$ is an exact multiple of $u$ we may of course replace the composite urns by $u$ urns each containing $n$ balls of which the proportion of red balls is $p_u$. When this is so, the variance of the distribution of the player's mean score, *i. e.* $V(M_{ba})$, is necessarily zero and $V_b = M(V_{ba})$. If the condition of replacement holds good $V_{ba}$ does *not* depend on the total number of balls in the urn and it is immaterial whether we define in one or other way stated above what condition we impose when we make the sampling process homogeneous. This is not so when there is *no replacement.* Equations (3. 6) and (3. 7) do not then define the alternative stated above, *viz.* sampling from a universe of $u$ urns each containing $n$ balls

and each having the same proportionate composition as the composite urn containing *un* balls. To that extent we may regard the criterion of explanation as ambiguous for the case of non-replacement.

This approach to the partition of variance in the Lexian universe clarifies a distinction concerning which the following remarks of Aitken [2] suggest a widely current misapprehension:

> The results which we have obtained for the Lexian and Coolidge schemes exhibit the variance as *resolved into separate components of variance.* The Bernoullian component may be called the random component, since it arises even when probability is constant, while the Lexian component may be called the *systematic* component, since it arises from the systematic alteration or variation of probability from one experiment to another. This resolution of variance has been called *analysis of variance.* It has been greatly extended by Professor R. A. Fisher, who has devised regular schemes of experimental arrangement involving many variates, by means of which not one but several systematic components of variance can be isolated from each other and from the random component.

The partition of variance defined by (3. 8, 3. 9) which is equivalent to the familiar formula for the difference between the variance of the distribution of sampling in the Lexian and Bernoullian universes derives its rationale from postulates totally different from those which justify the identification of explained variation with the so-called best estimate of the variance of the column means in a one-way classification of scores of a grid such as one which summarises a plot yield experiment. A Fisher score grid of this type specifies a single sample from a universe of scores and any parameters based thereon are therefore estimates. The Lexian score-frequency grid with which we here deal is a complete distribution summarising the results of extracting an indefinitely large number of all possible samples, and the appearance of the factor $(s-1)/s$ on the right hand side of (3. 9) has no connection with the problem of estimation. The equation itself is an exact description of how much variance arises from the circumstance of stratification in the entire universe of choice.

What circumstances do in fact justify a partition formally identical with (3. 8, 3. 9) in the type of sampling which the Fisher score grid embodies need not in fact concern us since Churchill Eisenhart has made them explicit in the communication already cited. It suffices to say that they are irrelevant to the particular assumptions which permit

---

[2] *Statistical Mathematics*, p. 54.

us to define a meaningful partition of variance in terms of causality in a Lexian set-up, and do not involve an important restriction which makes any such partition of variance in the latter unique. For the Lexian model universe is highly artificial inasmuch as each sub-universe is of equal size; and equations (3. 8, 3. 9) suffice to evaluate how much variance arises from the stratification of the universe if, and only if, our concern is with stratification subject to this limitation. Needless to say, few circumstances in practical statistics will arise when it holds good.

### 4. SAMPLE SIZE AS THE SOURCE OF VARIATION

We now consider the relation between a system in which one score ($x_a$) is the size of the sample the player takes from an urn of $n$ balls in which the fixed proportion of red balls is $p$. If $x_b$ is as before the actual score of the player:

$$M_{ba} = x_a \cdot p \quad ; \quad M_b = p \cdot E_a(x_a) = p \cdot M_a \ ,$$
$$\therefore M_{ba} - M_b = p(x_a - M_a).$$

Thus regression is linear with respect to the player's score on the A score, and

$$k_{ba} = p,$$
$$\therefore \ V(M_{ba}) = p^2 \cdot V_a \ . \tag{4.1}$$

If there is no replacement:

$$V_{ba} = \frac{x_a(n - x_a)}{n-1} pq = \frac{pq}{n-1} (nx_a - x_a^2),$$
$$\therefore M(V_{ba}) = \frac{npq}{n-1} E_a(x_a) - \frac{pq}{n-1} (V_a + M_a^2)$$
$$= \frac{pq}{n-1} M_a(n - M_a) - \frac{pq}{n-1} V_a \ . \tag{4.2}$$

From (4. 1) and (4. 2) we obtain:

$$V_b = M(V_{ba}) + V(M_{ba}) = \frac{pq}{n-1} M_a(n - M_a) - \frac{pq}{n-1} V_a + p^2 V_a$$
$$\therefore V_b = \frac{pq}{n-1} M_a(n - M_a) - \frac{p(1 - np)}{n-1} V_a \ . \tag{4.3}$$

We thus obtain:

$$r_{ab}^2 = \frac{V(M_{ba})}{V_b} = \frac{p(n-1) \cdot V_a}{qM_a(n - M_a) + (np - 1) V_a} \ . \tag{4.4}$$

When there is replacement

$$V_{ba} = x_a pq \quad ; \quad M(V_{ba}) = M_a pq;$$

$$\therefore V_b = M_a pq + p^2 V_a; \tag{4.5}$$

$$\therefore r_{ab}{}^2 = \frac{pV_a}{qM_a + pV_a} \ . \tag{4.6}$$

The last expression is of course the limiting form of (4.4) when $n$ is very large, so that $(n-1) \simeq n \simeq (n - M_a)$ and $(np - 1) \simeq np$.

In this set-up the variance of the player's score distribution depends on two circumstances, *viz.*, random sampling irrespective of sample size and variation with respect to sample size itself. There is a unique sense in which we can partition $V_b$ to exhibit how much of it is explicable in virtue of the distinctive peculiarity of the model, *viz.*, sample size variation, if we conceive the procedure as an indefinitely repeated sequence of sets of trials involving the same distribution of sample size per set. So conceived, the number of items chosen in any one such set remains the same if for every trial of $x_a$ items we substitute $M_a$, the mean sample size. This involves the restriction that $E_a(x_a)$ is an *integer*. If there is no replacement the residual (unexplained) variance is then:

$$V_U = \frac{M_a(n - M_a)}{n-1} pq \ . \tag{4.7}$$

Thence from (4.2) and (4.3):

$$M(V_{ba}) = V_U - \frac{pq}{n-1} V_a \ ,$$

$$V_b = V_U + V_E = V_U + \frac{p(np-1)V_a}{n-1} \ .$$

Since also $V_b = M(V_{ba}) + V(M_{ba})$, we thus derive:

$$V_U = M(V_{ba}) + \frac{q}{p(n-1)} \cdot V(M_{ba}), \tag{4.8}$$

$$V_E = \left[1 - \frac{q}{p(n-1)}\right] V(M_{ba}). \tag{4.9}$$

When $n$ is indefinitely large we derive the corresponding expressions for replacement:

$$V_U = M(V_{ba}) \quad \text{and} \quad V_E = V(M_{ba}). \tag{4.10}$$

Subject to the condition that there is replacement, the variance of

the player's mean score distribution therefore tallies exactly with the component of explanation, as is also true of the consequential relation of the player's score to that of the umpire in the models of Part I, though *not* of the concurrent relation between the scores of the two players in the same set-up. But we arrive at this conclusion by considerations which are in no way comparable with those we invoke to interpret a one-way partition of variance of the player's score distribution with respect to the contribution of the umpire in the umpire-bonus model.

A property of the umpire-bonus model in the domain of consequence is that there is always linear regression of the player's score on that of the umpire and the grid is homoscedastic in the same dimension. The variance of the player's score distribution for a fixed contribution of the umpire is thus constant, and the player's mean score is proportionate to the umpire's score. This is on all fours with one set of assumptions which entitle us to identify with a particular source of variation the variance of the means of one set of arrays in the *universe* of a score grid referable to one criterion of classification, the implication being likewise that the relation involved is consequential. Indeed it would appear that constant variance of one set of arrays is a sufficient condition for the identification of a relationship as consequential and for the validity of the partition of variance exhibited in equation (4. 10). On the other hand, homoscedasticity is not a necessary criterion of one or the other, for we have here derived (4. 10) for a set-up which is consequential though the grid is homoscedastic in neither dimension.

In this class of models, as for all others dealt with in this contribution with the exception of the special case discussed in § 6 below, regression is linear in only one dimension of the grid. A peculiarity of the Class B model recalls that of the binomial lottery model of Part II. When there is replacement, $V_{ba}$ like $M_{ba}$ is a linear function of $x_a$, the sample size; otherwise, *i. e.* without replacement, $V_{ba}$ is a quadratic function of $x_a$. The expressions for $r_{ab}$ in (4. 2) and (4. 4), like those of (3. 3) and (3. 4), only invoke parameters of the distribution of sample size, *i. e.* the source of variation. When the sample size distribution is rectangular within a set of trials and runs consecutively from $m$ to $t$ ($\leqq n$),

$$V_a = \frac{(t-m+1)^2-1}{12} \quad \text{and} \quad M_a = 1 + \frac{(t-m)}{2} .$$

If $m = 1$ we then derive for non-replacement in agreement with Example $2a$ of § 2 above:

$$r_{ab}^2 = \frac{p(n-1)(t-1)}{3q(2n-t-1) + (np-1)(t-1)} .$$

Subject to the same condition when there is replacement, and in agreement with Example $2b$ in § 2 above,

$$r_{ab}^2 = \frac{p(t-1)}{6q + p(t-1)} .$$

When the replacement condition holds good, the Class B model has a special interest inasmuch as it discloses the possibility of a peculiar type of *one*-way stratification consistent with exact correspondence between the universe values of $V(M_c)$ and $M(V_c)$ respectively with explained and residual components of variation. If we think of the latter in terms of "treatments" each applied to an otherwise identical distribution of scores, we are at liberty to postulate that:

(a) the effect of the entire system of treatments is not such as to alter the grand mean of all score values;

(b) each treatment has the effect of simultaneously changing the mean score and the variance of a complete column of score distributions in the same ratio;

(c) to each treatment evoking a $k$-fold increase of the mean column score, there corresponds a second treatment evoking a $k$-fold decrease.

Collectively we may thus express this simultaneous change of origin and scale of the original ($x$) score distribution, by denoting the final score corresponding to the $m$-th pair of treatments which are complementary in the sense implied by (c):

$$u_m = (1 + k_m)^{\frac{1}{2}}x + [(1 + k_m) - (1 + k_m)^{\frac{1}{2}}]M_x ,$$
$$w_m = (1 - k_m)^{\frac{1}{2}}x + [(1 - k_m) - (1 - k_m)^{\frac{1}{2}}]M_x .$$

The corresponding increments of the mean score are then:

$$E(u_m) - M_x = k_m \cdot M_x \quad ; \quad V(u_m) - V_x = k_m \cdot V_x ;$$
$$E(w_m) - M_x = -k_m \cdot M_x \quad ; \quad V(w_m) - V_x = -k_m \cdot V_x .$$

Likewise we have

$$u_m - E(u_m) = (1 + k_m)^{\frac{1}{2}}x - (1 + k_m)^{\frac{1}{2}} \cdot M_x \ ,$$

$$w_m - E(w_m) = (1 - k_m)^{\frac{1}{2}}x - (1 - k_m)^{\frac{1}{2}} \cdot M_x \ ,$$

$$E_{rc}[u_m - E(u_m)]^2 = (1 + k_m)E_{rc}(x^2) - (1 + k_m)M_x^2 = (1 + k_m)V_x \ ,$$

$$E_{rc}[w_m - E(w_m)]^2 = (1 - k_m)V_x \ .$$

If we have $r$ sets of $u$-scores $u_1, u_2 \cdots u_r$, etc., and $r$ corresponding sets of $w$-scores $w_1, w_2 \cdots w_r$ ,

$$M(V_c) = \frac{rV_x \cdot E_c(1 + k_m) + rV_x \cdot E_c(1 - k_m)}{2r} = V_x.$$

Evidently also

$$V(M_c) = M_x^2 E_c(k_m^2).$$

The last expression is the increase of variance due to the treatments.

The sub-universe score distributions of the Class B replacement model thus differ *inter se* in a way which is equivalent to simultaneous change of scale and origin. Needless to say, the necessary condition of this simultaneous change specified by (c) is highly artificial, but none the less suggestive inasmuch as it prompts us to explore a type of stratification which implies change of scale alone. If the effect of the treatment is to increase the variance of the column score distribution without changing the mean, the variance of the column means remains zero and any total increase of variance arises in virtue of an increase of the mean variance within columns. In the balance sheet of the composite universe, the reality of its stratification need not appear as a difference between total and residual variance, as is evident if we envisage the treatment as effecting change of scale without changing the mean value of the residual ($x$-) score distribution. We may then represent our new score system in the $m$-th column as

$$u_m = k_m^{\frac{1}{2}}x + (1 - k_m^{\frac{1}{2}})M_x \ .$$

In this case, we have

$$M(V_c) = E(k_m)V_x \ .$$

Thus

$$M(V_c) = V_x \text{ if } E_c(k_m) = 1.$$

The only circumstance to distinguish a stratified universe so conceived from the homogeneous parent universe would in fact be that it would not be homoscedastic with respect to the treatment criterion.

### 5. UNIT SAMPLE RECTANGULAR MODEL

Models of Class B discussed in the last section, subject to the replacement condition, are not without relevance to situations which arise in human biology, *e. g.*, where there is approximately linear regression of family size on some numerically specifiable characteristic of the social environment, such as income level. Expectation accomplishment models of Class C in § 1 designated as above are worthy of attention inasmuch as they illustrate the genesis of a coefficient of correlation possibly serviceable [3] as a summarising index of the influence of birth order on the incidence of rare congenital conditions if they occur singly in a sibship. If parity has no effect in the long run, expectation with respect to birth rank of affected individuals in a sibship of $s$ members will be equal for rank $1, 2 \cdot \cdot \cdot s$. A grid lay-out with family size as one set of border scores and birth rank of affected individuals as the alternate set will then exhibit a triangular contour; and the birth rank mean will increase by equal steps with respect to equally spaced family size. On the same assumption there will be a correlation between the two sets of scores; and the product-moment coefficient must be less than unity. To the extent that affected individuals crop up mostly among first births or towards the end of the family, the observed value of the coefficient will approach zero or unity respectively. Thus the discrepancy between the value of the product-moment coefficient ($r_0$) computed from observed data and its value ($r_{ab}$) computed in accordance with the null hypothesis of equal expectation is indicative of the extent to which birth rank influences the occurrence. It is therefore of some interest to explore the magnitude of $r_{ab}$ *vis à vis* the family size distribution.

Formally we have defined the Class C model in § 1 above, *i. e.*, the score $x_a$ is the number of cards in one of a set of packs from each of which the player selects a single card. He records the number of pips

---

[3] If $r_0$ is the product-moment coefficient for the *observed* bivariate distribution of affected individuals to each of which we assign one score ($x_a$) in virtue of family size and one score ($x_b$) in virtue of birth rank and $r_{ab}$ is the corresponding coefficient computed on the assumption that the expected birth rank of the individual has a rectangular distribution as in models of this class, we may define an index ($B_r$) which has the value zero when there is no birth rank effect with limits of $+1$ and $-1$ respectively, signifying that all effects are last-born or first-born, *viz.*

$$B_r = \frac{(r_0 - r_{ab})(2r_0 \cdot r_{ab} - r_0 - r_{ab} + 1)}{r_{ab}(1 - r_{ab})} .$$

as his own score $x_b$ referable to a particular pack the cards of which have $1, 2, 3 \cdots$ pips up to $s$, the size of the pack. No more than one card of a particular denomination is present in the pack, and hence the unit sample distribution with respect to any pack is rectangular. In the biological illustration $x_a$ and $x_b$ respectively correspond to family size and birth rank. The mean value of $x_b$ associated with $x_a$ is evidently $\frac{1}{2}(x_a + 1)$, so that

$$M_b = \tfrac{1}{2}E_a(x_a + 1) = \tfrac{1}{2}(M_a + 1),$$

and

$$M_{ba} - M_b = \tfrac{1}{2}(x_a - M_a).$$

Thus regression is necessarily linear in the B-dimension of the grid, irrespective of the distribution of card pack size and $k_{ba} = \frac{1}{2}$. Since $k_{ba}^2 \cdot V_a = V(M_{ba})$,

$$V(M_{ba}) = \tfrac{1}{4}V_a \ .$$

From the elementary property of the rectangular distribution,

$$V_{ba} = \frac{x_a^2 - 1}{12},$$

$$\therefore M(V_{ba}) = \frac{E_a(x_a^2)}{12} - \frac{1}{12} = \frac{V_a + M_a^2}{12} - \frac{1}{12}$$

$$\therefore M(V_{ba}) = \frac{V_a}{12} + \frac{M_a^2 - 1}{12} \ . \qquad (5.1)$$

$$\therefore V_b = \frac{V_a}{4} + \frac{V_a}{12} + \frac{M_a^2 - 1}{12}$$

$$= \frac{V_a}{3} + \frac{M_a^2 - 1}{12}, \qquad (5.2)$$

$$\therefore \frac{V(M_{ba})}{V_b} = r_{ab}^2 = \frac{3V_a}{4V_a + M_a^2 - 1} \ . \qquad (5.3)$$

The evaluation of $r_{ab}$ now depends solely on the distribution of card-pack- (in our biological illustration *family*-) size ($x_a$). Two cases admit of easy solution:

(a) *rectangular* distribution of family size from 1 to $N$:

$$M_a = \frac{N+1}{2} \quad ; \quad M_a^2 = \frac{(N+1)^2}{4} \quad ; \quad V_a = \frac{N^2 - 1}{12} \quad ;$$

$$\therefore r_{ab}^2 = \frac{3(N+1)}{7N + 13} \ . \qquad (5.4)$$

(b) *binomial* distribution of family size. We assume that consecutive values of $x_a$ from 1 to $N$ occur with frequencies defined by successive terms of the binomial $(q+p)^{N-1}$, so that:

$$M_a = (N-1)p + 1 \quad \text{and} \quad V_a = (N-1)pq,$$

$$\therefore V_b = \frac{(N-1)p}{12}[6 + (N-5)p],$$

$$\therefore r_{ab}^2 = \frac{3q}{6 + (N-5)p}\ . \tag{5.5}$$

Examples 3*a* and 3*b* in § 2 above illustrate the validity of the two formulae last cited.

The variance of the player's score in the unit sample rectangular model set-up admits of a unique partition in terms of the contribution attributable to the circumstance that the card packs are of unequal size in a set of trials only if we agree on a somewhat arbitrary prescription of how we propose to eliminate this source of variation. The problem admits of a simple solution if our prescription is to replace each individual card pack of a set of unit trials by one and the same pack of $M_a$ cards. It is however arguable that we might pool the $N$ packs of variable size in each $N$-fold set of trials and record the result of taking $N$ unit samples with replacement from the composite pack. Such a procedure admits of no singular solution unless we specify the distribution of card pack size, and hence of the unit sample distribution of the composite pack. With due recognition that the alternative procedure is highly arbitrary it is however of some interest to examine its implications. For the residual variance $V_u$ when we eliminate in this way the source of variation arising from the fact that the card packs are not all of the same size we may then write:

$$V = \frac{M_a^2 - 1}{12}.$$

Hence from (5.2), since $V(M_{ba}) = \frac{1}{4}V_a$:

$$V_b = \frac{V_a}{3} + V_u = \frac{4}{3}V(M_{ba}) + V_u,$$

$$\therefore V_E = V_b - V_u = \frac{4}{3}V(M_{ba}). \tag{5.6}$$

The admitted arbitrariness of this partition would scarcely be worthy of mention if it did not help to bring into focus the artificiality of any

unique partition of variance in the Lexian universe. For we have already seen that no unique solution of the problem of assigning a value to the component of variance identifiable with the circumstance of Lexian stratification is possible unless we impose the special restriction that each sub-universe is of equal size.

The models we have dealt with in this series are thus divisible into three categories:

(a) those which admit an unique partition of variance referable to the source of concomitant variation as is true of the models of Part I and for reasons stated of Class B models in this context;

(b) those which admit of no meaningful partition of variance, as is true of the models of Part II;

(c) those which admit of such a partition only if we impose a somewhat artificial condition relevant to the possibility of removing the source of concomitant variation or agree to adopt a somewhat arbitrary prescription for eliminating it.

### 6. LINEAR REGRESSION IN BOTH DIMENSIONS

With one notable exception, it is a common property of the models under discussion that regression is always linear in one dimension and in one dimension only. A special case, illustrated by Example 1*a* in § 2, arises when we sample without replacement from a stratified universe in which each of $(n + 1)$ sub-universes contains the same number of items if the unit trial expectation from one or other is $0, 1/n, 2/n \cdots 1$. When the conditions last stated hold good, our numerical example illustrates two properties then peculiar to the model:

(a) the distribution of the row border scores like that of the column border scores is rectangular;

(b) regression of the row scores on the column scores is linear as is generally true of regression of column scores on row scores for other models here dealt with.

Both these properties depend on relations among binomial coefficients which may not be too well known.[4] We consider the identity

$$(1 - t)^{-k}(1 - t)^{-m} = (1 - t)^{-k-m}. \qquad (6.1)$$

[4] For these and others, see, for example, Netto, *Lehrbuch der Combinatorik*, Chapter 13.

Expanding each term by the binomial theorem and equating the coefficients of corresponding powers of $t$, we get

$$\sum_{x=0}\binom{k+x-1}{x}\binom{m+r-x-1}{r-x}=\binom{k+m+r-1}{r}. \quad (6.2)$$

In particular, for $k=m=n$, we have

$$\sum_{x=0}\binom{n+x-1}{x}\binom{n+r-x-1}{r-x}=\binom{2n+r-1}{r}. \quad (6.3)$$

In the model illustrated in Example 1*a* of § 2, we have as the probability, $P(x;c)$ of drawing $x$ red and $s-x$ black balls from an urn containing $c$ red and $n-c$ black balls

$$P(x;c)=\binom{c}{x}\binom{n-c}{s-x}\Big/\binom{n}{s}. \quad (6.4)$$

To obtain the probability of getting $x$ red balls regardless of the urn we have to evaluate

$$P(x)=\sum_c P(x;c)P(c),$$

or since

$$P(c)=1/(n+1),$$

we need

$$P(x)=\frac{1}{n+1}\sum_c\binom{c}{x}\binom{n-c}{s-x}\Big/\binom{n}{s}, \quad (6.5)$$

where $c$ runs from 0 to $n$. It is easy to see, however, that the terms of the sum vanish except for

$$x\leqq c\leqq n-s+x;$$

and using (6.2) we obtain without difficulty

$$\sum_c\binom{c}{x}\binom{n-c}{s-x}=\binom{n+1}{s+1}. \quad (6.6)$$

From (6.5) and (6.6) we obtain at once

$$P(x)=\frac{1}{n+1}\binom{n+1}{s+1}\Big/\binom{n}{s}$$

$$=\frac{1}{s+1}, \quad (6.7)$$

showing that the row frequency distribution, like the column distribution, is rectangular.

The demonstration that the regression of column scores on row scores is linear is straightforward if we write

$$M_{cx} = \sum_{c=0} cP(c;x)$$

and make use of the fact that

$$P(c;x)P(x) = P(c,x) = P(x;c)P(c).$$

We have from (6.4) and (6.7), with a little manipulation,

$$M_{cx} = \sum_{c} c\binom{c}{x}\binom{n-c}{s-x}\Big/\binom{n+1}{s+1}; \tag{6.8}$$

and now we can write

$$c\binom{c}{x} = (x+1)\binom{c+1}{x+1} - \binom{c}{x},$$

whence, after some reduction involving the use of (6.2) we obtain

$$M_{cx} = \frac{(x+1)(n+2)}{s+2} - 1, \tag{6.9}$$

which we see is linear in $x$, with regression coefficient

$$\frac{n+2}{s+2}.$$

## 7. SUMMARY

The primary aim of the examination of models dealt with in this and in preceding contributions to this series is, as the title suggests, semantic rather than mathematical. The outcome brings into focus the need to distinguish between two procedures commonly subsumed under the title *analysis of variance*. One is the comparison of estimates derived from different ways of classifying an assemblage of data with a view to assessing the credentials of a null hypothesis which postulates a homogeneous universe. The other which arises only if we have reason to reject the null hypothesis, has as its aim to present a balance sheet of variance exhibiting what fractions of the total variance of a set of scores are severally referable to sources of variation inherent in the classificatory framework. The logical implications of the first pro-

cedure and its mathematical credentials are justifiable without recourse to the assumption that there is any single formula for the second or indeed that its undertaking is realisable.

Only the second procedure engages our attention in this context, and only insofar as any balance sheet of the components of variance based on information a sample supplies derives its rationale from the implicit assumption of a specifiable partition of variance within the stratified universe from which it comes.[5] It is therefore necessary to emphasise that the models of this series of communications are models of universe stratification in contradistinction to models of sampling procedures. As such they disclose situations in which:

(a) a meaningful and unique partition of variance in terms of the source of variation is possible;

(b) a meaningful partition of variance on such terms is impossible unless we invoke arbitrary definitions of what the removal of a contributory source of variation implies;

(c) no such partition of variance is meaningful or possible, since the removal of the constraint in virtue of which a correlation between two sets of scores arises does not in fact alter the total variance of either distribution.

It is logically permissible to distinguish as *consequence, concurrence* and *constraint,* three types of relationship which the algebra of correlation takes within its scope, but such a trichotomy cuts across the dividing lines between (a), (b) and (c) above, except insofar as models of Part II specified by (c) are models of constraint. Situations in which correlation arises in virtue of a consequential relationship may fall within the scope of (b) or of (a) above; and if it is indeed meaningful to specify a moiety of the variance of the consequent score as *explained variation* the condition of linear regression does not suffice to justify its identification with the square of the product-moment index. In the domain of concurrence, the only model (Part I of this series) we have studied does admit of a meaningful specification of explained variation. Regardless of whether regression is or is not linear the

[5] More explicitly, the partition (Eisenhart's Model II) of a 2-way stratified system specified by $\sigma_r^2 + \sigma^2 e/c$ as the expected sample value of $V(M_c)r/(r-1)$ and by $\sigma_c^2 + \sigma^2 e/r$ as the expected sample value of $V(M_r)c/(c-1)$ refers to sampling from a universe in which the overall values of $V(M_r)$ and $V(M_c)$ are respectively $\sigma_r^2$ and $\sigma_c^2$.

fraction of the variance of either score distribution attributable to the source of concomitant variation is then the ratio of the covariance of the joint distribution thereto.

There is indeed nothing sacrosanct about the additive relation between the variance of the distribution of the means of the columns or rows and the mean of the variance of the distribution of scores in these columns or rows *vis à vis* the total variance of row or column border scores. It is, as Churchill Eisenhart has emphasised, a tautology of a grid regardless of any statistical properties inherent in its lay-out. It is not a universal formula for disclosing the causal nexus therein. By the same token, there is nothing sacrosanct about the correlation ratio as an index of explanation. The circumstance of linear regression does not suffice to establish its title as such, since it is possible to devise many models which exhibit linear regression consistent with a different partition of variance with respect to its sources. What balance sheet of causality is permissible depends upon the nature of the model whose properties the grid itself summarises. The bearing of this conclusion on the cognate procedure called analysis of covariance scarcely calls for comment.

# THE URINARY 17-KETOSTEROID OUTPUT OF NAVAL SUBMARINE ENLISTED CANDIDATES DURING TWO STRESSFUL SITUATIONS*

BY LT. COMDR. ELLSWORTH B. COOK (MSC) USN
*Department of Pharmacology, Tufts Medical School*
*Boston, Massachusetts*

AND

PROFESSOR ROBERT J. WHERRY
*Department of Psychology, Ohio State University*
*Columbus, Ohio*

## INTRODUCTION

THIS paper discusses the urinary 17-ketosteroid[1] excretion of a group of naval enlisted submarine candidates during two selected situations. This area study comprised one section of an investigation

* The investigations reported herein were conducted at the U. S. N. Medical Research Laboratory, U. S. N. Submarine Base, New London, Connecticut, under Bureau of Medicine and Surgery Research Project NM-003-017. Opinions expressed are those of the authors and are not to be construed as necessarily reflecting the views or the endorsement of the Navy Department.

[1] Some 20 hormones have been extracted from the adrenal cortex; all belong to the class of substances known as steroids—four carbon ring structures with relatively simple side chains and with molecular weights around 300. Steroids are excreted in the urine primarily in the form of 17-ketosteroids, that is, as steroids with an oxygen atom attached to a particular carbon atom (#17) in one of the rings. The 17-ketosteroids are end-products of steroid hormone metabolism. They are considered indicative of adrenal cortex activity because of the following observed facts (1):

(a) In Addison's disease in which the adrenal cortex becomes non-functional, the decline in 17-ketosteroid output corresponds to the severity of the disease.
(b) In certain cancerous growths of the adrenal cortex there is a marked outpouring of 17-ketosteroids.
(c) They are excreted in very nearly normal amounts by castrated males and ovariectomized females, indicating the minor role of the gonads in their production.

of the possible value of a variety of measures in the selection of submarine personnel (2).

Accepting the hypothesis that one's emotional stability is related to the responsiveness of the adrenal cortex, the problem of this study may be stated thus:

(A) Is the taking of special psychological tests sufficiently stressful to submarine enlisted candidates to be reflected in significant changes in their 17-ketosteroid output?

(B) Is the undergoing of routine training at the Escape Training Tank at the New London Submarine Base sufficiently stressful to candidates to be reflected in significant changes in their 17-ketosteroid output?

*Background.* Selye has studied the train of events which follows injury in animals, regardless of the nature of the damaging agent (exposure to cold, traumatic injuries, spinal shock, acute infections, and intoxications with various drugs) (3, 4, 5). Among the most prominent of the functional changes he observed upon exposure of animals to stress was an enlargement of the adrenal cortex with increased corticoid-hormone secretion. This response (the adaptation syndrome or alarm reaction) he interprets as indicative of the increased demand by the body for cortical hormones (6). When stress is placed upon an organism, the normal output of the adrenal glands is inadequate; this discrepancy between supply and demand results in a temporary state of relative adrenal insufficiency, and the adrenal cortex becomes activated to meet the increased demand.

This was confirmed for humans by Browne *et al.* (7, 8). Studies on hospital patients by Forbes and her colleagues (9, 10) also indicate that following a noxious stimulus of any kind, there is an accelerated action of the adrenal cortex as judged by a rise in 17-ketosteroid output.

Hoagland, Pincus and their associates have found an enhancement of the output of 17-ketosteroid substances by normal persons in response to a wide variety of stresses—exposure to heat, to cold, to anoxia, to fatiguing pursuit meter tasks, and to difficult examinations (11, 12, 13, 14, 15). In contrast, the output of these hormonal metabolites during stress was greatly reduced or absent in the case of psychotics. This failure of fundamental alarm response mechanisms in the latter is noteworthy since psychotics are persons who have broken under the stresses of daily life.

The differences in vulnerability to stress thus revealed among medically healthy individuals was extended to an investigation of service conditions. Studies conducted for the Air Forces by the Worcester Foundation for Experimental Biology indicated that the stresses of flying were reflected in the urinary excretion from the adrenal cortex (16). Moreover, the 17-ketosteroid output of a group of pilots was found to correlate positively and significantly with their superior officer's rating of their fatigability. Such physiological reaction to stress appears, then, to hold promise for the meaningful classification of individuals who are faced with the exacting emotional strains of modern aerial and submarine life.

## EXPERIMENTAL PROCEDURE

A total of 120 subjects were employed in the study. They were selected at random from the general group of enlisted candidates undergoing routine processing for Submarine School. Two groups of six subjects each were tested weekly during an intensive three-day experimental period (2). One stress situation was scheduled for the morning of the second and third experimental days. For the first thirteen groups the psychological stress situation was given initially; the order was reversed for the remaining seven groups (2).

I. *The psychological stress situation.* This consisted of the administration of the short group form of the Minnesota Multiphasic Personality Inventory (hereinafter referred to as the MMPI) and the Officers Classification Test (hereinafter referred to as the OCT).

The MMPI is a psychometric instrument developed (17) to assist psychologists and other personnel workers who must deal with personality difficulties among more nearly normal persons rather than with the obvious abnormalities with which psychiatry is concerned.

The OCT is a battery of aptitude tests designed to measure verbal ability, mechanical comprehension, mechanical and electrical information, mechanical ability and spatial perception (18).

The customary time allotment was permitted for completion of the MMPI, and the personal nature of the questions asked was relied upon to produce stress. The OCT was regarded as beyond the general ability of most enlisted candidates, and, in an attempt to increase stress effects, less than half the period usually given for its completion was allowed. Additional tension was provided during both tests by periodically

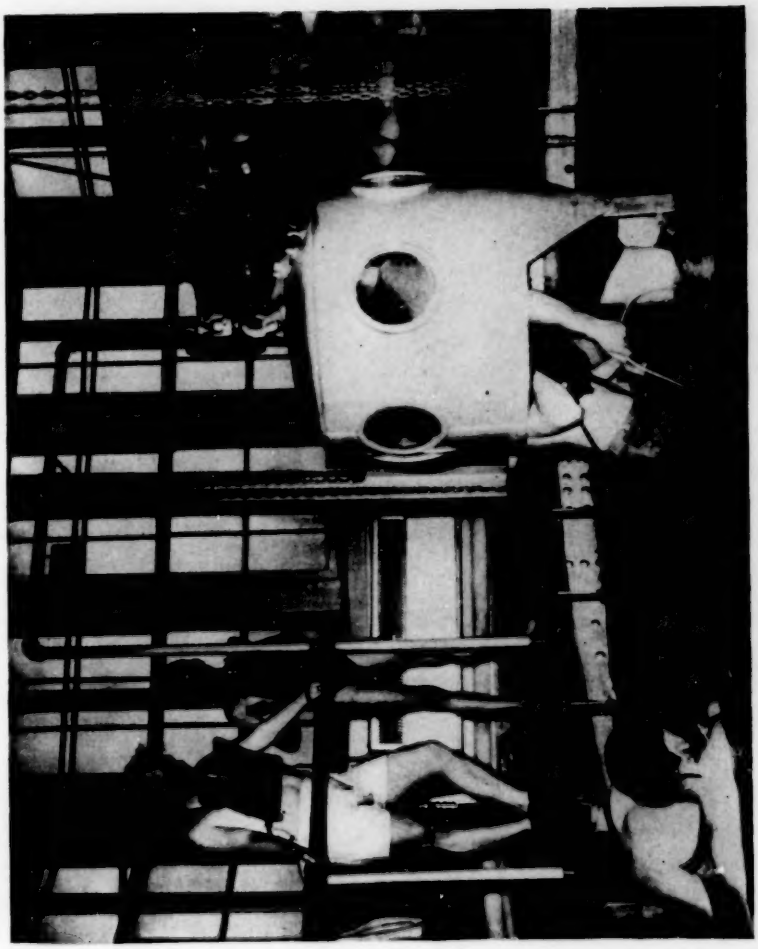

FIG. 1. INTERIOR OF SUBMARINE ESCAPE TANK SHOWING TRAINING BELL.
Subjects are wearing Momson lungs.

announcing the sections of each which should have been completed by that particular time interval.

II. *The tank stress situation.* This consisted of the routine training procedures undergone by submarine enlisted candidates at the Submarine Escape Training Tank. This device, intended to acquaint personnel with the method of escaping from a submerged submarine which is unable to rise to the surface, is a tower containing a column of fresh water 25 feet in diameter and 100 feet deep. A training bell permits an ascent from any desired depth, and hatches or locks, are located at depths of 18, 50 and 100 feet. The interior of the tank is shown in Fig. 1.

As a preliminary check on their ability to stand pressure, the subjects entered a dry recompression chamber where they were given 50 lbs. (3.4 atmospheres) in from 6-10 minutes. Each man was then examined by a qualified otologist for ear damage (19). Men who passed this inspection were instructed in the use of the Momson lung, an inflatable device containing a chemical carbon dioxide absorbent which enables the user to breathe freely while underwater. Then each man made two escapes from the training bell at 12 feet and two escapes from each of the 18 and 50-foot locks.

During the course of each stress situation, four urine samples were taken by two chief hospital corpsmen:

1. The basal—the first sample collected upon arising 0600 (approx.)
2. The pre-stress—taken 15 minutes before stress began 0800 "
3. The stress—taken 15 minutes after the stress ended 1100 "
4. The post-stress—taken 2 hours after the stress ended 1245 "

These specimens were forwarded daily in special sealed containers for analysis by the Worcester Foundation for Experimental Biology (20, 21). An aliquot of 500 cc. was sent for each subject for each sample, and the total volume excreted was recorded on the specimen container. Two or three drops of toluene were added to each jar as a preservative. The amount of creatinine in the urine was determined as a check on the samples taken, and androgens were extracted for possible points of interest related to changing 17-ketosteroid output.

### STATISTICAL ANALYSIS AND RESULTS

Data on 85 of the 120 subjects were utilized in the statistical analysis. The remaining men were dropped because of difficulty in the dry recompression chamber, failure to complete tank training, or because their urine samples were incomplete (breakage of specimen containers in transit, inadequate creatinine content, and so forth). The variables selected for analysis are listed in Table 1, together with their means and standard deviations.

TABLE 1

*Variables selected for statistical analysis, with their means and standard deviations*

| | VARIABLE NUMBER AND DESCRIPTION | MEAN | S. D. |
|---|---|---|---|
| 01 | 17-ketosteroid-psych. basal sample in mgm./hr. | 0.247 | 0.130 |
| 02 | " " " pre-stress " " " | 0.356 | 0.223 |
| 03 | " " " stress " " " | 0.294 | 0.191 |
| 04 | " " " post-stress " " " | 0.344 | 0.180 |
| 05 | " " tank basal " " " | 0.233 | 0.115 |
| 06 | " " " pre-stress " " " | 0.399 | 0.249 |
| 07 | " " " stress " " " | 0.346 | 0.222 |
| 08 | " " " post-stress " " " | 0.338 | 0.269 |
| 09 | Age in years | 18.47 | 1.31 |
| 10 | Creatinine-psych. basal sample in gm./24 hr. | 1.784 | 0.650 |
| 11 | " " pre-stress " " " " | 1.797 | 0.463 |
| 12 | " " stress " " " " | 2.017 | 0.545 |
| 13 | " " post-stress " " " " | 2.046 | 0.504 |
| 14 | " tank basal " " " " | 1.458 | 0.356 |
| 15 | " " pre-stress " " " " | 1.793 | 0.493 |
| 16 | " " stress " " " " | 2.004 | 0.552 |
| 17 | " " post-stress " " " " | 1.985 | 0.561 |
| 18 | 17-ketosteroid mgm./hr. (av. var. 01 and 05) | 0.240 | 0.093 |
| 19 | Androgen mgm./hr. | 0.119 | 0.070 |
| 20 | Total testicular volume in cu. cm. | 43.6 | 14.8 |

It has been established that there is a diurnal ketosteroid rhythm in normal healthy individuals going about their daily routine (22, 23, 24, 25). This basic ketosteroid excretion is relatively constant for the individual, and has been compared to the resting blood pressure as characteristic of the individual (10). Typically, the pattern exhibits

a low output during sleep and rises rapidly an average of 60 per cent upon awaking. It then falls in an essentially linear fashion during the course of a routine day, reaching minimal levels again at night as indicated in Fig. 2.

Thus, had the four urine samples taken in this study been obtained under ordinary conditions, the greatest output of 17-ketosteroid sub-

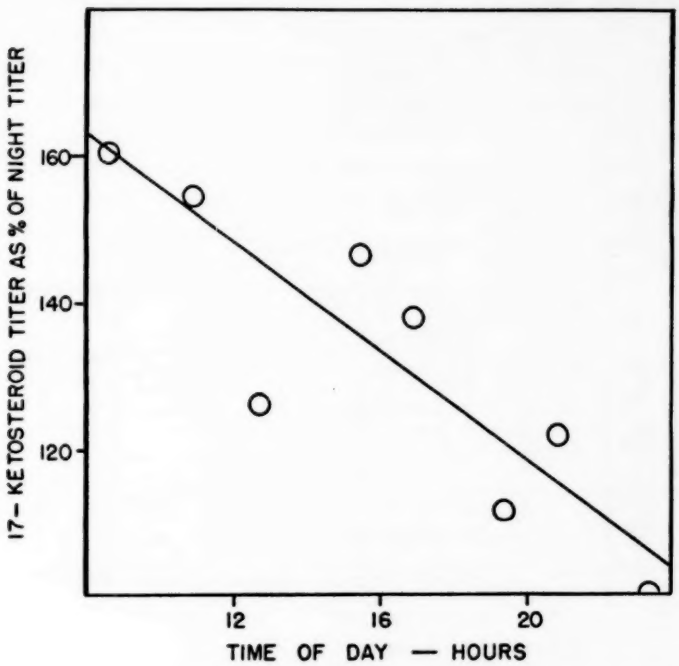


FIG. 2. OUTPUT OF 17-KETOSTEROID AS A FUNCTION OF TIME OF DAY
From Pincus and Hoagland data (11).

stances would be evidenced in the basal sample (0600) and would decrease in amount with each subsequent specimen.

A consideration of the normal daily output is necessary, then, before meaning can be attached to the increase exhibited under stress.

Since subjects were available only during the three-day experimental period (2), there was no opportunity to determine their normal excretion inasmuch as none of the three days could be regarded as usual or routine for the men. To establish an equation for the normal output of the 85 subjects who ranged in age from 18-24, Pincus and Hoagland

data (11) for a similar age group [2] were employed (see Fig. 2). This indicated that the trend line for normal output between the hours of 0600-1300 is a straight line, $y = 195 - 3.83x$, where $y$ equals ketosteroid output and $x$ equals time of day in hours.

The trend lines for 17-ketosteroid output of the subjects before, during, and after the two situations are compared with the trend line of normal output in Fig. 3. The basal sample is represented by 100,

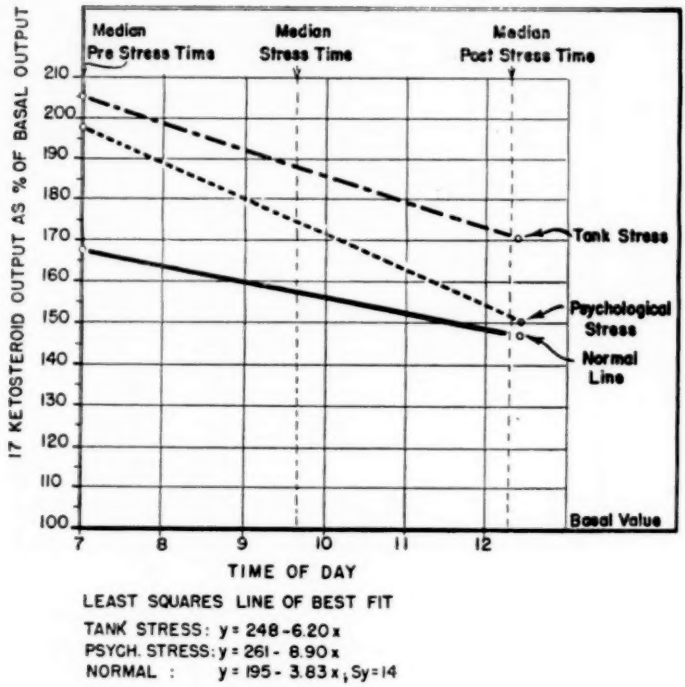


FIG. 3. COMPARISON OF NORMAL AND STRESS SITUATION 17-KETOSTEROID MEAN OUTPUT

Normal output from Pincus and Hoagland (11).

and the other specimens as percentage increase over basal. This figure indicates (a) that both situations resulted in greatly enhanced output of 17-ketosteroid substances, (b) that for both situations, the greatest increase over basal output occurred in the pre-stress sample, and (c)

[2] 97 specimens of 7 male Clark University students between the ages of 18 and 25 were used to establish the normal diurnal variation (11).

that tank stress resulted in a greater excretion of 17-ketosteroid substances than did the psychological stress where recovery to the normal was accomplished at a faster rate.

Table 2 indicates the various categories of individual reactions as determined by ketosteroid samples, (a) in the basal to pre-stress period, and (b) in the pre-stress to stress period. Thus, reading across the top line of Table 2, it can be seen that of the 52 subjects who increased their output during both situations in the basal to pre-stress period, 9 men continued to increase for both situations in the pre-stress to stress period as well, 5 continued to increase for the psychological but decreased for the tank situation, 11 continued to increase for the tank but decreased for the psychological situation, and 27 decreased for both situations in the pre-stress to stress period compared with the excretion in their basal to pre-stress samples.

TABLE 2

*Trend (increases and decreases) in 17-ketosteroid output during the (a) basal to pre-stress, and (b) pre-stress to stress periods of the two situations*

| FROM BASAL TO PRE-STRESS | | FROM PRE-STRESS TO STRESS | | | | TOTAL | |
|---|---|---|---|---|---|---|---|
| | | Inc. psych. Inc. tank | Inc. psych. Dec. tank | Inc. tank Dec. psych. | Dec. psych. Dec. tank | No. | % |
| Inc. psych. Inc. tank | | 9 | 5 | 11 | 27 | 52 | 61 |
| Inc. psych. Dec. tank | | 3 | 1 | 6 | 1 | 11 | 13 |
| Inc. tank Dec. psych. | | 2 | 10 | 0 | 5 | 17 | 20 |
| Dec. psych. Dec. tank | | 4 | 0 | 0 | 1 | 5 | 6 |
| Total | No. | 18 | 16 | 17 | 34 | 85 | |
| | % | 21 | 19 | 20 | 40 | | 100 |

Summarizing this table, 94 per cent of the men exhibited an increased ketosteroid output during at least one situation, 74 per cent increased during the psychological, 81 per cent during the tank, and 61 per cent during both situations. Only 6 per cent of the subjects exhibited a

decreased output during the two phases which Table 2 examines for both situations.

Of the 61 per cent who increased their ketosteroid output during both situations, approximately half (27 out of 52) exhibited the greatest increase in the basal to pre-stress period, and a slightly lower output during actual stress itself. This conservation of ketosteroids during the short stress periods to which the men were exposed is in accord with the results of a study on test pilots (15) where a compensatory hypoexcretion of ketosteroid substances was found for periods involving brief portions of actual flight time.

Table 3 carries the breakdown of individual subject reactions still further, and examines the trend of 17-ketosteroid output from stress to post-stress for each cell entry of Table 2. Thus, reading across the top line of Table 3, it can be seen that of the 9 subjects (first cell entry of Table 2) who increased their output for the two situations during both (a) the basal to pre-stress period, and (b) the pre-stress to stress period also, 3 men continued to increase output for both situations during the stress to post-stress period as well, 1 continued to increase during this phase of the tank situation but decreased for the psychological situation, and 5 men decreased for both situations in their post-stress ketosteroid samples.

There was a wide variation evidenced in individual responses to the two situations. "Stress output" ranged from 21.2% to 950.0% of the basal sample in the case of the psychological situation, and from 14.0% to 2628.6% in the case of the tank situation. Several typical ketosteroid responses are shown graphically in Fig. 4. The mean increase of the various samples over the basal specimen is presented in Table 4, with the standard deviation for each.

In general, subject response may be summed up by the statement that there was a greatly enhanced output of 17-ketosteroid substances during both situations, and a tendency to feel more apprehension while anticipating the stress than when actually undergoing it. It should prove interesting to relate these reactions to criterion data in a later paper in view of the Hoagland finding that of the men who exhibited the fundamental alarm mechanisms during a pursuit meter task, those best conserving their steroids under the stress were the best performers (13).

The normal daily output of creatinine is between 1.5 and 2.0 grams for adult males. The function is considered extraordinarily constant for

TABLE 3

*Breakdown of cell entries in Table 2 to show trend of 17-ketosteroid output from stress to post-stress periods*

| FROM PRE-STRESS TO STRESS | FROM STRESS TO POST-STRESS | | | | TOTAL |
|---|---|---|---|---|---|
| | Inc. psych. and tank | Inc. psych. Dec. tank | Inc. tank Dec. psych. | Dec. psych. and tank | |
| *A. Basal to pre-stress—increase on psych. and tank situations* | | | | | |
| Inc. psych. and tank | 3 | 0 | 1 | 5 | 9 |
| Inc. psych., dec. tank | 1 | 4 | 0 | 0 | 5 |
| Inc. tank, dec. psych. | 0 | 6 | 5 | 0 | 11 |
| Dec. psych. and tank | 12 | 8 | 7 | 0 | 27 |
| Total | 16 | 18 | 13 | 5 | 52 |
| *B. Basal to pre-stress—increase on psych., decrease on tank* | | | | | |
| Inc. psych. and tank | 0 | 0 | 2 | 1 | 3 |
| Inc. psych., dec. tank | 0 | 1 | 0 | 0 | 1 |
| Inc. tank, dec. psych. | 1 | 2 | 1 | 2 | 6 |
| Dec. psych. and tank | 1 | 0 | 0 | 0 | 1 |
| Total | 2 | 3 | 3 | 3 | 11 |
| *C. Basal to pre-stress—increase on tank, decrease on psych.* | | | | | |
| Inc. psych. and tank | 0 | 1 | 0 | 1 | 2 |
| Inc. psych., dec. tank | 5 | 0 | 2 | 3 | 10 |
| Inc. tank, dec. psych. | 0 | 0 | 0 | 0 | 0 |
| Dec. psych. and tank | 4 | 1 | 0 | 0 | 5 |
| Total | 9 | 2 | 2 | 4 | 17 |
| *D. Basal to pre-stress—decrease on psych. and tank* | | | | | |
| Inc. psych. and tank | 0 | 0 | 2 | 2 | 4 |
| Inc. psych., dec. tank | 0 | 0 | 0 | 0 | 0 |
| Inc. tank, dec. psych. | 0 | 0 | 0 | 0 | 0 |
| Dec. psych. and tank | 0 | 0 | 0 | 1 | 1 |
| Total | 0 | 0 | 2 | 3 | 5 |

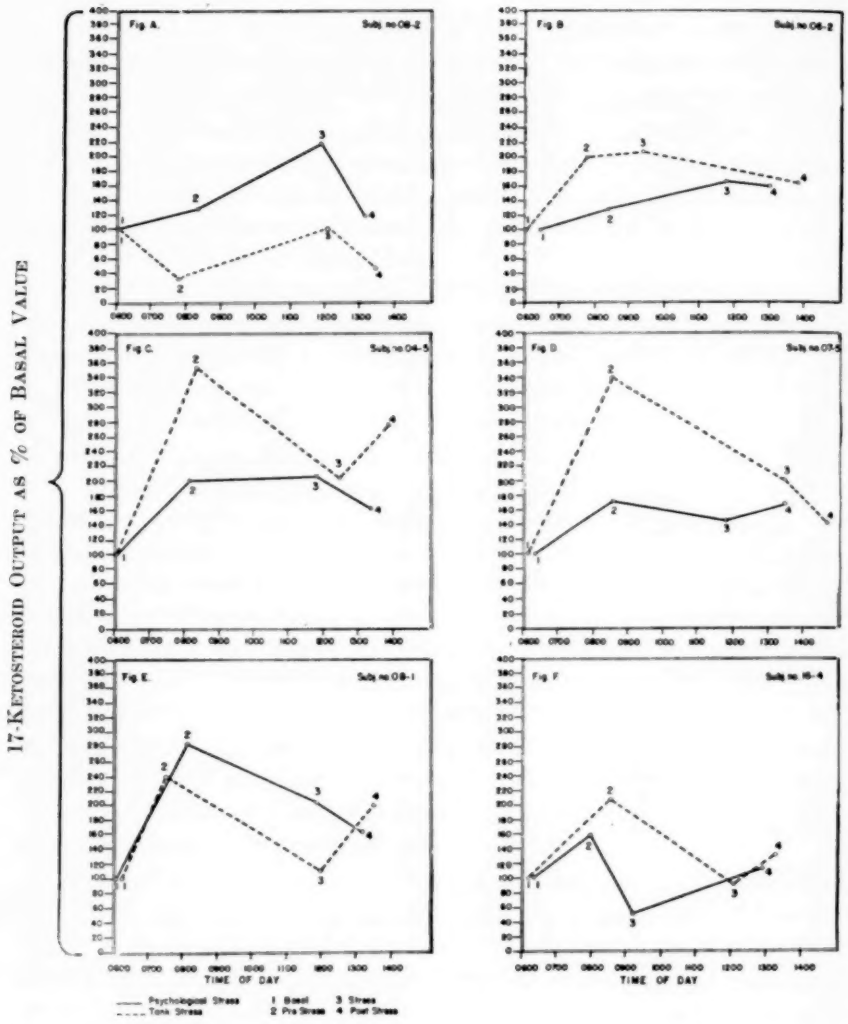


Fig. 4. Some Typical 17-Ketosteroid Response Graphs Plotted as Percentage of Basal Value, at the Actual Time Sample Was Obtained

TABLE 4

*Mean 17-ketosteroid output during the stress situation expressed as percentage increase over basal samples*

| SAMPLE | PSYCHOLOGICAL STRESS SITUATION | | TANK STRESS SITUATION | |
|---|---|---|---|---|
| | Mean | S. D. | Mean | S. D. |
| Pre-stress | 201.2% | 108.4 | 209.4% | 286.4 |
| Stress | 158.7% | 114.4 | 181.2% | 262.8 |
| Post-stress | 155.2% | 92.4 | 176.5% | 305.5 |

an individual (26), that is to say, A with an output of 1.5 and B with an output of 2.0 are both considered to be within the normal range, but A's daily output would consistently be 1.5 and that of B tend to be consistently 2.0. Keeping this in mind, let us examine the mean trend of creatinine production during the two stress situations. It is obvious from Fig. 5 that the mean output of creatinine rose during both situations. Changes were not quite statistically significant. However, they did occur and were sufficiently great in the case of enough subjects to influence the mean trend. This suggests the possibility that there may be individual differences in the ability to maintain a constant creatinine production during stress. No explanation is offered for the differences in creatinine production exhibited during psychological as compared to the tank situation, and further study appears warranted.

The product moment correlation coefficients obtained between each pair of the 20 variables are presented in Table 5. A modified (27) Thurstone Group Centroid (28) factor analysis [3] yielded nine orthogonal

[3] Factor analysis is the statistical technique which attempts to explain a large number of intercorrelations in terms of a smaller number of factors or underlying bases of association. Results are presented in terms of factor loadings; these represent the correlation between a given variable and one of the factors isolated. They may be positive or negative in sign, depending on the nature of the relationship with the particular variable involved. The factor loading squared gives the percentage of variance of a given measurement which may be explained or predicted by the factor in question. In the discussion which follows, a loading of .20 or higher is considered significant. Designations assigned the factors isolated represent the interpretative judgment of the authors, and the reader is invited to consider and suggest alternate names.

factors with factor loadings on the 20 variables as indicated in Table 6. These loadings reproduced the correlations very closely, leaving residuals (Table 5) with a mean of 0.003 and a standard deviation of 0.059.

The factors isolated by the analysis fall into three groups: ketosteroid, androgen and creatinine factors.

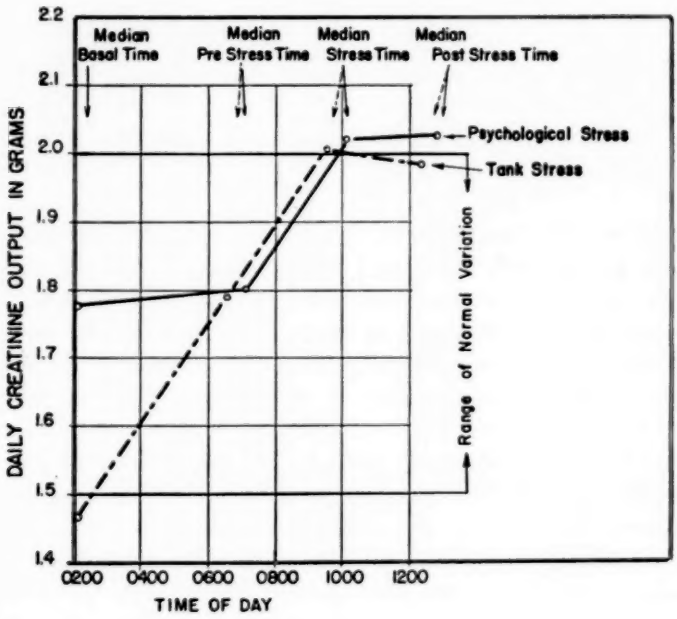


FIG. 5. COMPARISON OF MEAN CREATININE PRODUCTION FOR TANK AND PSYCHOLOGICAL STRESS SITUATIONS

I. *Ketosteroid factors.* Factor A has moderately high and significant loadings on all measures of ketosteroid output (variables 01, 02, 03, 04, 05, 06, 07, 08, and 18) regardless of time or of the stressful situation involved. Accordingly this factor is considered to represent general ketosteroid production and has been labelled *basic ketosteroid output.* The highest loading (.81) is found on variable 18, the average of the output for the two basal periods (variables 01 and 05). This indicates the higher reliability and validity of this composite measure as contrasted with any single measurement.

TABLE 5

*Intercorrelations and residuals of variables in ketosteroid analysis*

Population = 85; significance levels: $P = .05, |r| \geq .21$; $P = .01, |r| \geq .28$

Intercorrelations

| Test | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | | .42 | .40 | .52 | .20 | .55 | .25 | .31 | .05 | -.18 | .25 | .11 | .13 | .13 | .14 | .08 | .29 | .74 | .40 | .14 |
| 02 | .05 | | .39 | .41 | .13 | .40 | .34 | .35 | .03 | .05 | .08 | .23 | -.08 | .21 | .06 | .09 | .21 | .29 | .12 | .08 |
| 03 | .07 | .05 | | .38 | .09 | .28 | .41 | .40 | -.01 | -.08 | .08 | -.07 | .07 | .00 | -.01 | .09 | .15 | .33 | .25 | -.03 |
| 04 | -.03 | .02 | .04 | | .17 | .39 | .24 | .39 | .14 | .00 | -.06 | .08 | -.04 | -.07 | .08 | .08 | .11 | .48 | .24 | .06 |
| 05 | -.06 | .07 | -.08 | -.06 | | .21 | .25 | .13 | .08 | -.09 | .14 | .18 | .18 | .03 | -.12 | -.05 | .14 | .73 | .53 | .04 |
| 06 | .10 | -.03 | -.09 | -.02 | .05 | | .31 | .39 | .15 | -.01 | .13 | .18 | .10 | .18 | .12 | .02 | .20 | .46 | .32 | .12 |
| 07 | .00 | .00 | .09 | -.07 | -.03 | -.08 | | .34 | .14 | .00 | .03 | -.04 | -.04 | -.04 | -.28 | -.18 | .07 | .37 | .14 | .04 |
| 08 | -.05 | -.02 | .05 | .01 | -.02 | -.01 | -.01 | | -.01 | -.23 | .10 | -.10 | -.10 | .10 | -.04 | -.04 | .05 | .35 | .20 | .09 |
| 09 | -.04 | -.03 | -.04 | .00 | -.01 | .07 | .01 | -.06 | | .04 | .04 | .09 | -.05 | -.04 | .05 | -.02 | .03 | .09 | -.10 | .03 |
| 10 | .05 | -.07 | .00 | .08 | -.01 | .02 | .01 | -.04 | .01 | | -.29 | .40 | .10 | .26 | .06 | .17 | .16 | -.22 | -.17 | .01 |
| 11 | -.03 | .03 | .00 | -.03 | .03 | -.03 | .00 | .00 | .03 | -.08 | | .20 | .51 | .20 | .24 | .13 | .29 | .28 | .08 | -.07 |
| 12 | -.03 | .08 | -.08 | .08 | .03 | .08 | -.07 | -.04 | .07 | .02 | -.02 | | .33 | .32 | .16 | .30 | .22 | .12 | .05 | -.11 |
| 13 | -.06 | -.09 | .09 | .07 | .02 | .04 | -.01 | -.02 | -.06 | .02 | -.05 | -.09 | | .23 | .19 | .28 | .41 | .22 | -.01 | -.12 |
| 14 | .07 | .09 | -.02 | -.09 | .06 | .10 | .01 | .10 | -.05 | .04 | .04 | .07 | .03 | | .25 | .28 | .28 | .08 | .04 | .03 |
| 15 | .03 | -.03 | .00 | .03 | .00 | .06 | -.07 | -.01 | .07 | -.09 | .01 | -.05 | -.08 | -.02 | | .40 | .36 | .02 | .07 | -.15 |
| 16 | .01 | -.03 | .08 | .00 | .02 | -.05 | -.02 | .01 | .03 | -.09 | .00 | .05 | .03 | .03 | -.05 | | .33 | -.04 | -.01 | -.11 |
| 17 | .07 | .01 | .03 | .02 | .03 | -.01 | .02 | -.06 | .03 | .07 | -.07 | -.08 | .03 | .04 | .05 | .04 | | .28 | .12 | -.14 |
| 18 | .01 | .00 | .01 | -.02 | .03 | .08 | .01 | .00 | -.07 | -.03 | .04 | -.08 | .03 | .05 | .08 | .04 | .09 | | .59 | .07 |
| 19 | .04 | -.07 | -.01 | -.02 | .07 | .09 | .00 | -.07 | .08 | .00 | .01 | -.02 | .00 | .02 | .07 | -.06 | -.04 | .07 | | .04 |
| 20 | .05 | .05 | -.07 | -.03 | .00 | .08 | -.01 | .03 | .00 | .05 | .00 | -.06 | .01 | .05 | -.07 | -.02 | -.08 | -.05 | -.03 | |

Residuals

TABLE 6

*Final factor loadings*

| | VARIABLE NO. AND DESCRIPTION | A | B | C | D | E | F | G | H | I | $h^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 17 KS psych.-basal | .54 | .13 | .72 | —.07 | .06 | .16 | .15 | —.06 | —.01 | .89 |
| 02 | pre-stress | .41 | .43 | .10 | —.19 | —.03 | .09 | .06 | .18 | .16 | .47 |
| 03 | stress | .37 | .39 | .09 | —.03 | .05 | .00 | .08 | —.04 | .04 | .31 |
| 04 | post-stress | .50 | .33 | .35 | .01 | —.07 | .09 | —.18 | —.07 | .14 | .55 |
| 05 | 17 KS tank-basal | .47 | .01 | .05 | .72 | .13 | .02 | .09 | .01 | —.04 | .77 |
| 06 | pre-stress | .42 | .48 | .16 | —.08 | —.04 | .15 | .14 | .01 | .00 | .48 |
| 07 | stress | .43 | .46 | —.02 | .11 | —.17 | —.15 | .12 | .10 | —.15 | .51 |
| 08 | post-stress | .47 | .38 | .05 | —.11 | .00 | —.07 | .10 | —.16 | .10 | .43 |
| 09 | Age | .15 | —.01 | .06 | .10 | —.33 | .04 | —.04 | .01 | .00 | .19 |
| 10 | Creatinine psych.-basal | —.12 | .06 | —.19 | .01 | —.09 | .28 | —.19 | .71 | .11 | .69 |
| 11 | pre-stress | .08 | —.07 | .12 | .01 | —.02 | .27 | .71 | —.17 | .02 | .63 |
| 12 | stress | .11 | —.09 | .05 | .08 | .04 | .32 | .28 | .52 | .09 | .49 |
| 13 | post-stress | —.10 | —.11 | .19 | .19 | —.02 | .30 | .68 | .20 | .08 | .69 |
| 14 | Creatinine tank-basal | .08 | —.05 | —.07 | —.11 | —.01 | .38 | .12 | .16 | .17 | .24 |
| 15 | pre-stress | —.13 | —.01 | .10 | —.10 | .04 | .53 | .10 | —.02 | .41 | .50 |
| 16 | stress | —.17 | .11 | .09 | —.02 | .13 | .49 | .04 | .13 | .37 | .46 |
| 17 | post-stress | .13 | .09 | .03 | .01 | .07 | .38 | .37 | .08 | .21 | .36 |
| 18 | 17 ketosteroid mgm./hr. | .81 | —.05 | .40 | .38 | .05 | .05 | .15 | .00 | —.01 | .99 |
| 19 | Androgen mgm./hr. | .49 | .10 | .05 | .19 | .70 | .04 | .03 | —.09 | .07 | .79 |
| 20 | Total testicular volume | .18 | —.05 | .02 | —.05 | .00 | —.02 | —.13 | —.03 | —.09 | .06 |

Factor labels:

A—Basic KS output
B—Stress KS output
C—Specific-duplicate reporting
D—Specific-duplicate reporting
E—Androgen production
F—Creatinine production
G—Perseveration in creatinine production
H—Creatinine-psych. specific
I—Creatinine-tank specific
$h^2$—Communality

Of considerable physiological interest is the relatively high loading (.49) of androgen output (variable 19) on this general ketosteroid factor, suggesting that the two are associated in some manner.

Ketosteroid output is unrelated to age within the limits (18-24) of this population, or to total testicular volume in any significant manner, and none of the creatinine measures have significant projections on it.

Factor B has significant loadings on the pre-stress, stress and post-stress samples of both situations, and is considered to represent *production of ketosteroid during stress.* The loadings indicate that there are wide individual differences in such increase (the means for variables 02, 03, 04 and for 06, 07, 08 are higher than those for variables 01 and 05). For both situations, loadings are highest during the pre-stress period (.43 for psychological and .48 for tank) and slightly lower (.39 for psychological and .46 for tank) during stress itself.

The fact that all measures high on this factor also have significant loadings on the basic ketosteroid production factor A as well, indicates the necessity for using the normal ketosteroid excretion as a corrective measure in estimating stress output. This may be accomplished by reporting either (a) a direct difference, or (b) a ratio—in effect, a difference of the logs.

The isolation of factor B supports the hypothesis which led to this study—there *is* a fundamental mechanism operative during stress which results in a significant increase in 17-ketosteroid output, and the situations to which the subjects were exposed were sufficiently stressful to induce this hormonal response.

Factors C and D represent spurious overlap due to the duplicate reporting of results. Factor C is concerned with the overlap of variable 18 with variable 01, and factor D with the overlap of variables 18 and 05. Thus, the two factors merely indicate that variable 18 was obtained by averaging variables 01 and 05. Neither has any physiological significance.

II. *Androgen factor.* Factor E has a significant positive loading (.70) only on variable 19 (androgen mgm./hr.) and is accordingly identified as an *androgen production factor.* The negative loading on age (—.39 for variable 09) clearly indicates a decrease of androgen output with age. Since the average age of the subjects in this study was only about 18½ years (221.6 months) and the standard deviation only a little over

I—Creatinine-tank specific
$h^2$—Communality
D—Specific-duplicate reporting
E—Androgen production

one year (15.7 months), this negative correlation is considered definitely significant. Obviously, any such androgen measures should be corrected for age in some fashion. The factor seems of doubtful importance as a predictive device, but should prove interesting in comparison with masculinity measures from other areas of the total study (2).

III. *Creatinine factors.* Factor F has significant positive loadings on all eight creatinine measures (variables 10 through 17) regardless of time or stress situation. It is therefore regarded as a *general creatinine production* factor.

No other measures have significant projections on this factor; apparently creatinine production is unrelated to (a) ketosteroid production or (b) androgen production.

TABLE 7

*Perseveration of creatinine production indicated by factor G*

| VARIABLE | MEAN (in gms.) | DIFFERENCE OF SUCCESSIVE MEANS | FACTOR LOADINGS |
|---|---|---|---|
| 10 | 1.78 | | |
| 11 | 1.80 | .02 | .71 |
| 12 | 2.02 | .22 | .28 |
| 13 | 2.05 | .03 | .68 |
| 14 | 1.46 | .59 | .12 |
| 15 | 1.79 | .33 | .10 |
| 16 | 2.00 | .21 | .04 |
| 17 | 1.99 | .01 | .37 |

Factor G has significant loadings on the creatinine measure of both post-stress specimens, as well as on the pre-stress and stress samples of the psychological situation. Loadings tend to be highest when the differences between two creatinine measures are smallest. This is indicated in Table 7 which compares factor loadings and the differences in successive means. Thus, factor G appears to measure the degree of constancy of creatinine in each specimen and accordingly it is labelled *perseveration in creatinine production.* It is regarded as some reaction common to the two post-stress periods, possibly associated with the delay in return of creatinine to normal levels after stress.

While the interpretation is not clearcut, the emergence of this independent factor G in addition to the general creatinine production

factor, F, holds promise for the detection of individual tolerance to stress as indicated by the differences in the degree of perseveration (resistance to change) of creatinine production. As suggested in connection with Fig. 5, this area is considered worthy of additional study.

Factor H has significant positive loadings on the creatinine measures of the basal and stress samples of the psychological stress situation, and a just barely significant loading (.20) on the post-stress sample of this situation. Factor I has significant loadings on the pre-stress, stress and post-stress tank measurements. These two factors are considered specifics for the psychological (factor H) and tank (factor I) situations. They are probably best regarded as residual situational elements in the stress series after factors F and G are subtracted. Regarded thus, they appear to be of no further importance.

One purpose of each area study of the type reported herein is to select single or composite measures, of the truly basic factors isolated, for relational analysis to significant components for inclusion in the final relational matrix. Factors A, B, E, F and G were chosen to represent this area.

## SUGGESTED ADDITIONAL EXPERIMENTATION

It is suggested that there is a fruitful field for investigation of 17-ketosteroid output as it relates to personnel response to various submarine operations, snorkelling[4] in particular.

Several experimental designs could be worked out to estimate such items as the stress effects of continued submergence upon crewmen, the differences in individual vulnerability to monotony and confinement, crew tolerance limits for snorkel operations and so forth.

Technical problems incidental to such studies would have to be solved (for example, how best to handle analysis of urines during prolonged snorkel sessions) but the potentialities of this type of investigation are rich and varied.

## CONCLUSIONS

1. There are basic production rates of both 17-ketosteroid substances and of creatinine which affect all measurements within each series.

[4] The snorkel is a device which permits submarines to recharge their batteries without surfacing.

2. Individual differences in the production rates of 17-ketosteroids and of creatinines are unrelated.
3. The output of androgen is positively related to the general level of 17-ketosteroid output, and negatively related to age.
4. The 17-ketosteroid output of subjects of this study increased significantly over basal samples during both stressful situations. The highest increase for both situations was during the pre-stress period, and a greater output was evidenced during the tank than during the psychological stress situation.
5. The amount of 17-ketosteroid increase during stress is considered a reliable measure of individual differences. The multiple correlation coefficient of .58 between the combination of the two stress ketosteroid measures and the stress factor, B, indicates that a single measure would be an inadequate basis for judgments concerning individual differences.
6. There is evidence for reliable individual differences in the degree of perseveration (resistance to change) with which creatinines are produced.
7. It is suggested that additional experimentation be undertaken at some future date to determine ketosteroid excretion during prolonged periods of snorkelling.

## REFERENCES

1. PINCUS, G. Adrenocortical hormones in human urine. *J. Clin. Endo.*, 3, no. 2, December 1943.
2. COOK, E. B., and R. J. WHERRY. A study of the interrelationships of psychological and physiological measures on naval submarine enlisted candidates. I. History, experimental design and statistical treatment of data. U. S. N. Medical Research Laboratory, U. S. N. Submarine Base, New London, Conn., Report No. 1, Bureau of Medicine and Surgery Research Project NM-003-017, March 9, 1949.
3. SELYE, H. The General Adaptation Syndrome and the Diseases of Adaptation. *Montreal*: Textbook of Endocrinology, *Acta Endrocrinologica*, Montreal University, Canada, 1947.
4. SELYE, H. Thymus and adrenals in the response of the organism to injuries and intoxication. *Brit. J. Exper. Path.*, 17, 1936.

5. SELYE, H. *et al.* On the therapeutic value of adrenal cortical hormones in traumatic shock and allied conditions. *J. Canad. Med. Assoc.*, 43: 1, 1940.

6. SELYE, H. Studies on adaptation. *Endo.*, 21, 1937.

7. BROWNE, J. S. I. The response of the adrenal cortex to damage. AAAS Hormone Conference, Mont Tremblant, Quebec, Sept. 20, 1944.

8. STEPHENSON, J. A. F., V. SCHENKER and J. S. I. BROWNE. The 17-ketosteroid excretion in damage and convalescence. *Endo.*, 35, September 1944.

9. FORBES, A. C. The 17-ketosteroid excretion in stress, preliminary report. *New York*: Josiah Macy Foundation, June 1942.

10. FORBES, A. C. *et al.* The effect of trauma and disease in the urinary 17-ketosteroid excretion in men. *J. Clin. Endo.*, 7, 1947.

11. PINCUS, G. and H. HOAGLAND. Steroid excretion and the stress of flying. *J. Av. Med.*, August 1943.

12. FREEMAN, W., G. PINCUS and E. D. GLOVER. The excretion of neutral urinary steroids in stress. *Endo.*, 35, September 1944.

13. HOAGLAND, H. Adventures in biological engineering. *Science*, July 28, 1944.

14. HOAGLAND, H. Enzyme kinetics and the dynamics of behavior. *J. Compar. and Physiol. Psych.*, 40, no. 3, June 1947.

15. HOAGLAND, H. The human adrenal cortex in relation to stressful activities. *J. Av. Med.*, 18, no. 5, October 1947.

16. PINCUS, G. and H. HOAGLAND. Effects of administered pregnenolone on fatiguing psychomotor performance. *J. Av. Med.*, April 1944.

17. HATHAWAY, S. R. and J. C. MCKINLEY. Manual of the Minnesota Multiphasic Personality Inventory. University of Minnesota Press, 1943.

18. U. S. NAVY. Officers classification test. Form X-1, Nav. Pers. 16680, 1944 (R).

19. SHILLING, C. W., H. C. HAINES, J. D. HARRIS and W. J. KELLY. The prevention and treatment of aerotitis media. *U. S. N. Med. Bulletin*, October 1945.

20. PINCUS, G. and W. H. PEARLMAN. Fractionization of neutral urinary steroids. *Endo.*, 29, no. 3, September 1941.

21. PINCUS, G. The analysis of human urines for steroid substances. *J. Clin. Endo.*, 5, no. 7, September 1945.

22. WERNER, S. C. The daily variation in 17-ketosteroid excretion of men and women. *J. Clin. Endo.*, 1, December 1941.

23. WOOSTER, H. A biometric study of total neutral 17-ketosteroid excretion in the normal adult male. *J. Clin. Endo.*, 4, September 1943.

24. TALBOT, N. B., A. M. BUTLER *et al.* Excretion of 17-ketosteroid by normal and abnormal children. *Am. J. Dis. Child.*, 65, 1943.

25. PINCUS, G. A diurnal rhythm in the excretion of urinary ketosteroids by young men. *J. Clin. Endo.*, 3, no. 4, April 1943.

26. BEST, C. H. and N. B. TAYLOR. Physiological Basis of Medical Practices, 3rd ed. *Baltimore*: Williams and Wilkins, 1943.

27. U. S. ARMY. Wherry-Brogden-Gaylord method of factor analysis. Personnel Research Section, Adjutant General's Office, Department of the Army, Washington, D. C. (unpublished).

28. THURSTONE, L. L. Multiple Factor Analysis. *Chicago*: University of Chicago Press, 1947.

# ON A GENERALIZATION OF THE BEHRENS-FISHER PROBLEM

BY JOHN E. WALSH
*The RAND Corporation*

## 1. *Summary*

THIS paper considers a generalization of the Behrens-Fisher problem which appears to be approximated by many practical situations. A solution is presented for the generalized situation and some efficiency properties of this solution are investigated.

2. *Introduction.* The Behrens-Fisher problem is concerned with testing the difference of the medians (means) of two normal populations when the ratio of the variances of the populations is unknown. Let one normal population have median $\mu$ while the other has median $\nu$. Then $\mu - \nu$ is tested on the basis of a sample (say of size $m$) from the population with median $\mu$ and an independent sample (say of size $n$) from the population with median $\nu$.

In some practical situations, however, it may only be known (approximately) that the $m + n$ observations are independent and that a specified $m$ of them are from continuous symmetrical populations with common median $\mu$ while the remaining $n$ are from continuous symmetrical populations with common median $\nu$. This is the generalization of the Behrens-Fisher problem which is considered in this paper. Then the results developed for the Behrens-Fisher problem are not necessarily applicable for testing $\mu - \nu$. This paper presents some tests for $\mu - \nu$ which are valid for the generalized situation. The method used also yields confidence intervals for $\mu - \nu$. It is to be emphasized that no two observations are necessarily drawn from the same population or from populations which have any common relationships other than those specified in the definition of the generalized problem.

In cases where the symmetry requirement is not satisfied, it is some-

times possible to obtain a transformed set of observations which do satisfy the generalized conditions by appropriate monotonic changes of variable. A special situation in which the generalized conditions are satisfied is that where $m$ of the observations are a sample from one continuous symmetrical population while the remaining $n$ observations are an independent sample from another continuous symmetrical population.

To obtain an idea of how much "information" is lost by using the tests developed for the generalized Behrens-Fisher problem when one actually has two independent samples from normal populations, two types of power efficiencies are computed for the generalized tests:

(a) Power efficiency with respect to Scheffé's "best" $t$-test solution when the ratio of variances is unknown (see [1]);

(b) Power efficiency with respect to the most powerful solution when the ratio of variances is known.

Although a "best" solution has been obtained for the Behrens-Fisher problem (see [2]), it was not computationally feasible to obtain the power efficiency of the generalized tests with respect to this "best" solution. Instead, upper and lower bounds were obtained for this efficiency. Case (a) represents an upper bound while case (b) represents a lower bound. For the particular situations analyzed in this paper, it is believed that the true power efficiency of a generalized test (compared with the "best" solution [2]) will be nearer the value for case (a) than the value for case (b).

The basic definition of what is meant by the power efficiency of a significance test is given in [3]. An intuitive discussion of the meaning of power efficiency is presented in [4]. For the cases considered, a power efficiency of $100v\%$ has the interpretation that the "best" test (*i. e.*, the test with respect to which the power efficiency is being found) based on $vm$ and $vn$ sample values has approximately the same power function as the given test (same significance level) based on $m$ and $n$ sample values.

All tests considered compare $\mu - \nu$ with a given hypothetical value $\phi_0$. Only one-sided and symmetrical tests are investigated. For the power efficiency analysis it is sufficient to limit consideration to one-sided tests of $\mu - \nu < \phi_0$. The power efficiencies of the corresponding one-sided tests of $\mu - \nu > \phi_0$ and symmetrical tests of $\mu - \nu \neq \phi_0$ can be obtained in the manner outlined in [3]. For simplicity of presen-

tation, it will be assumed that $n \leq m$. No generality is lost by this assumption.

3. *Outline of generalized solution.* The procedure used to obtain tests and confidence intervals which are valid for the generalized situation is as follows: Let $m = rn + s$, $(0 \leq s < n)$. Draw the $m$ observations from the populations with median $\mu$ as $s$ sets of $r+1$ observations and $n - s$ sets of $r$ observations. Number these sets according to the order in which they were drawn (*i. e.*, $1, \cdots, n$). Form the mean of the observations for each set and give each mean the same numbering as the set from which it was formed. Draw the $n$ observations from the populations with median $\nu$. Number these observations according to the order in which they were drawn and subtract the observation with number $i$ from the mean with the same number $(i = 1, \cdots, n)$. The resulting $n$ dummy observations are easily seen to be independent and from continuous symmetrical populations with median $\mu - \nu$. Thus the results of [3] and [5] can be applied to these observations to obtain tests and confidence intervals for $\mu - \nu$.

Let the values of the $n$ dummy observations, arranged in increasing order of magnitude, be represented by the bold face numbers $\mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{n}$. Then $\mathbf{1}$ denotes the value of the smallest dummy observation while $\boldsymbol{n}$ represents the value of the largest dummy observation, etc. Table 1 contains a list of some one-sided and symmetrical tests of $\mu - \nu$ for $n \leq 15$. These tests were obtained from those listed in [5, Tables 1 and 2].

The results of Table 1 can be modified to furnish one-sided and symmetrical confidence intervals for $\mu - \nu$. A one-sided significance test of the form

$$\textit{Accept } \mu - \nu < \phi_0 \textit{ if } g[\mathbf{1}, \cdots, \boldsymbol{n}] < \phi_0$$

with significance level $\alpha$ yields the one-sided confidence interval

$$(-\infty, \quad g[\mathbf{1}, \cdots, \boldsymbol{n}])$$

for $\mu - \nu$ with confidence coefficient $1 - \alpha$; *i. e.*,

$$\Pr\{\mu - \nu < g[\mathbf{1}, \cdots, \boldsymbol{n}]\} = 1 - \alpha.$$

Similarly, a one-sided test of the form

$$\textit{Accept } \mu - \nu > \phi_0 \textit{ if } h[\mathbf{1}, \cdots, \boldsymbol{n}] > \phi_0$$

TABLE 1

*Some one-sided and symmetrical significance tests for $n \leqq 15$*

Bold face numbers represent order statistics of the dummy observations described in the text.

| SIGNIFICANCE LEVEL (approximate) | | | TESTS | |
|---|---|---|---|---|
| | | | Symmetrical: Accept $\mu - \nu \neq \phi_0$ if either | |
| $n$ | One-sided % | Symmetrical % | One-sided Accept $\mu - \nu < \phi_0$ if given statistic $< \phi_0$ | One-sided Accept $\mu - \nu > \phi_0$ if given statistic $> \phi_0$ |
| 4 | 6.2 | 12.5 | (**4**) | (**1**) |
| | 5.0 | 10.0 | 1.055(**4**)—.055(**1**) | 1.055(**1**)—.055(**4**) |
| 5 | 6.2 | 12.5 | (**4**+**5**)/2 | (**1**+**2**)/2 |
| | 5.0 | 10.0 | .63(**5**)+.37(**4**) | .63(**1**)+.37(**2**) |
| | 3.1 | 6.2 | (**5**) | (**1**) |
| | 2.5 | 5.0 | 1.02(**5**)—.02(**1**) | 1.02(**1**)—.02(**5**) |
| 6 | 4.7 | 9.4 | max[**5**,(**4**+**6**)/2] | min[**2**,(**1**+**3**)/2] |
| | 3.1 | 6.2 | (**5**+**6**)/2 | (**1**+**2**)/2 |
| | 2.5 | 5.0 | .63(**6**)+.37(**5**) | .63(**1**)+.37(**2**) |
| | 1.6 | 3.1 | (**6**) | (**1**) |
| | 1.0 | 2.0 | 1.06(**6**)—.06(**1**) | 1.06(**1**)—.06(**6**) |
| 7 | 5.5 | 10.9 | max[**5**,(**4**+**7**)/2] | min[**3**,(**1**+**4**)/2] |
| | 2.3 | 4.7 | max[**6**,(**5**+**7**)/2] | min[**2**,(**1**+**3**)/2] |
| | 1.6 | 3.1 | (**6**+**7**)/2 | (**1**+**2**)/2 |
| | 1.0 | 2.0 | .785(**7**)+.215(**6**) | .785(**1**)+.215(**2**) |
| | 0.8 | 1.6 | (**7**) | (**1**) |
| | 0.5 | 1.0 | 1.05(**7**)—.05(**1**) | 1.05(**1**)—.05(**7**) |
| 8 | 4.3 | 8.6 | max[**6**,(**4**+**8**)/2] | min[**3**,(**1**+**5**)/2] |
| | 2.7 | 5.5 | max[**6**,(**5**+**8**)/2] | min[**3**,(**1**+**4**)/2] |
| | 1.2 | 2.3 | max[**7**,(**6**+**8**)/2] | min[**2**,(**1**+**3**)/2] |
| | 1.0 | 2.0 | max[**7**,{.5(**8**)+.28(**6**)+.22(**7**)}] | min[**2**,{.5(**1**)+.28(**3**)+.22(**2**)}] |
| | 0.8 | 1.6 | (**7**+**8**)/2 | (**1**+**2**)/2 |
| | 0.5 | 1.0 | .785(**8**)+.215(**7**) | .785(**1**)+.215(**2**) |
| | 0.4 | 0.8 | (**8**) | (**1**) |
| 9 | 5.1 | 10.2 | max[**6**,(**4**+**9**)/2] | min[**4**,(**1**+**6**)/2] |
| | 2.2 | 4.3 | max[**7**,(**5**+**9**)/2] | min[**3**,(**1**+**5**)/2] |
| | 1.0 | 2.0 | max[**8**,(**5**+**9**)/2] | min[**2**,(**1**+**5**)/2] |
| | 0.6 | 1.2 | max[**8**,(**7**+**9**)/2] | min[**2**,(**1**+**3**)/2] |
| | 0.5 | 1.0 | max[**8**,{.5(**9**)+.28(**7**)+.22(**8**)}] | min[**2**,{.5(**1**)+.28(**3**)+.22(**2**)}] |
| | 0.4 | 0.8 | (**8**+**9**)/2 | (**1**+**2**)/2 |

TABLE 1 (*continued*)

| $n$ | SIGNIFICANCE LEVEL (approximate) One-sided % | SIGNIFICANCE LEVEL (approximate) Symmetrical % | TESTS: Symmetrical: Accept $\mu - \nu \neq \phi_0$ if either: One-sided Accept $\mu - \nu < \phi_0$ if given statistic $< \phi_0$ | TESTS: Symmetrical: Accept $\mu - \nu \neq \phi_0$ if either: One-sided Accept $\mu - \nu > \phi_0$ if given statistic $> \phi_0$ |
|---|---|---|---|---|
| 10 | 5.6 | 11.1 | max[**6**,(**4**+**10**)/2] | min[**5**,(**1**+**7**)/2] |
| | 2.5 | 5.1 | max[**7**,(**5**+**10**)/2] | min[**4**,(**1**+**6**)/2] |
| | 1.1 | 2.1 | max[**8**,(**6**+**10**)/2] | min[**3**,(**1**+**5**)/2] |
| | 0.5 | 1.0 | max[**9**,(**6**+**10**)/2] | min[**2**,(**1**+**5**)/2] |
| 11 | 4.8 | 9.7 | max[**7**,(**4**+**11**)/2] | min[**5**,(**1**+**8**)/2] |
| | 2.8 | 5.6 | max[**7**,(**5**+**11**)/2] | min[**5**,(**1**+**7**)/2] |
| | 1.1 | 2.1 | max[(**6**+**11**)/2,(**8**+**9**)/2] | min[(**1**+**6**)/2,(**3**+**4**)/2] |
| | 0.5 | 1.1 | max[**9**,(**7**+**11**)/2] | min[**3**,(**1**+**5**)/2] |
| 12 | 4.7 | 9.4 | max[(**4**+**12**)/2,(**5**+**11**)/2] | min[(**1**+**9**)/2,(**2**+**8**)/2] |
| | 2.4 | 4.8 | max[**8**,(**5**+**12**)/2] | min[**5**,(**1**+**8**)/2] |
| | 1.0 | 2.0 | max[**9**,(**6**+**12**)/2] | min[**4**,(**1**+**7**)/2] |
| | 0.5 | 1.1 | max[(**7**+**12**)/2,(**9**+**10**)/2] | min[(**1**+**6**)/2,(**3**+**4**)/2] |
| 13 | 4.7 | 9.4 | max[(**4**+**13**)/2,(**5**+**12**)/2] | min[(**1**+**10**)/2,(**2**+**9**)/2] |
| | 2.3 | 4.7 | max[(**5**+**13**)/2,(**6**+**12**)/2] | min[(**1**+**9**)/2,(**2**+**8**)/2] |
| | 1.0 | 2.0 | max[(**6**+**13**)/2,(**9**+**10**)/2] | min[(**1**+**8**)/2,(**4**+**5**)/2] |
| | 0.5 | 1.0 | max[**10**,(**7**+**13**)/2] | min[**4**,(**1**+**7**)/2] |
| 14 | 4.7 | 9.4 | max[(**4**+**14**)/2,(**5**+**13**)/2] | min[(**1**+**11**)/2,(**2**+**10**)/2] |
| | 2.3 | 4.7 | max[(**5**+**14**)/2,(**6**+**13**)/2] | min[(**1**+**10**)/2,(**2**+**9**)/2] |
| | 1.0 | 2.0 | max[**10**,(**6**+**14**)/2] | min[**5**,(**1**+**9**)/2] |
| | 0.5 | 1.0 | max[(**7**+**14**)/2,(**10**+**11**)/2] | min[(**1**+**8**)/2,(**4**+**5**)/2] |
| 15 | 4.7 | 9.4 | max[(**4**+**15**)/2,(**5**+**14**)/2] | min[(**1**+**12**)/2,(**2**+**11**)/2] |
| | 2.3 | 4.7 | max[(**5**+**15**)/2,(**6**+**14**)/2] | min[(**1**+**11**)/2,(**2**+**10**)/2] |
| | 1.0 | 2.0 | max[(**6**+**15**)/2,(**10**+**11**)/2] | min[(**1**+**10**)/2,(**5**+**6**)/2] |
| | 0.5 | 1.0 | max[**11**,(**7**+**15**)/2] | min[**5**,(**1**+**9**)/2] |

with significance level $\alpha$ furnishes the one-sided confidence interval

$$(h[\mathbf{1},\cdots,\mathbf{n}],\infty)$$

with confidence coefficient $1-\alpha$. Finally, a symmetrical test of the form

$$\textit{Accept } \mu-\nu\neq\phi_0 \textit{ if either } g[\mathbf{1},\cdots,\mathbf{n}]<\phi_0 \textit{ or } h[\mathbf{1},\cdots,\mathbf{n}]>\phi_0$$

with significance level $2\alpha$ yields the symmetrical confidence interval

$$(h[\mathbf{1},\cdots,\mathbf{n}],\quad g[\mathbf{1},\cdots,\mathbf{n}])$$

with confidence coefficient $1-2\alpha$.

As an example of obtaining a one-sided confidence interval from a one-sided test of Table 1, let $n=7$ and consider the test

$$\textit{Accept } \mu-\nu<\phi_0 \textit{ if } .785(\mathbf{7})+.215(\mathbf{6})<\phi_0 .$$

Then

$$[-\infty,\ .785(\mathbf{7})+.215(\mathbf{6})]$$

is a one-sided confidence interval for $\mu-\nu$ with confidence coefficient approximately equal to .99 (or 99 per cent). As an example of obtaining a symmetrical confidence interval, let $n=6$ and consider:

$$\textit{Accept } \mu-\nu\neq\phi_0 \textit{ if either } \max[\mathbf{5},(\mathbf{4}+\mathbf{6})/2]<\phi_0$$
$$\textit{or } \min[\mathbf{2},(\mathbf{1}+\mathbf{3})/2]>\phi_0 .$$

Then

$$(\min[\mathbf{2},(\mathbf{1}+\mathbf{3})/2],\ \max[\mathbf{5},(\mathbf{4}+\mathbf{6})/2])$$

is a symmetrical confidence interval for $\mu-\nu$ with confidence coefficient .906 (90.6 per cent).

If the order in which the observations were drawn is not available, the sets of $m$ and $n$ observations can be combined to form the $n$ dummy observations in an unbiased manner by use of an independent randomization process.

This solution to the generalized Behrens-Fisher problem has the disadvantage that it is not invariant with respect to permutations within the set of $m$ observations or within the set of $n$ observations. It has the advantage, however, that it is valid under very general conditions.

4. *Example of application.* To demonstrate how the generalized solution to the Behrens-Fisher problem might be applied, let us consider a numerical example. The given data consist of a sample of size 6

from one continuous symmetrical population and a sample of size 13 from another continuous symmetrical population. Let the values of the sample of size 6, sequentially ordered according to the time of drawing, be

1.724, 2.612, 1.623, 5.334, 1.864, 3.414,

while the values for the sample of size 13 (also ordered according to time of drawing) are

—.494, 1.048, .346, .637, 2.176, —1.185, .972, 1.210, 2.647,
.398, .846, .654, .522.

It is desired to test whether the median of the first population minus the median of the second population exceeds unity at the 1.6% significance level; also to obtain a symmetrical confidence interval for this difference of medians which has a 95% confidence coefficient.

For this case, $n = 6$, $m = 13$, $s = 1$, $r = 2$, $\nu =$ median of first population, $\mu =$ median of second population. The 13 sample values are divided into one set of 3 observations and five sets of two observations. The first set consists of the values —.494, 1.048, .346; the second set of the values .637, 2.176; · · · ; the sixth set of the values .654, .522. By forming the six means for these sets, one obtains the values (in correct order),

.300, 1.407, —.107, 1.929, .622, .588.

Subtracting the corresponding values (in time ordering) for the sample of size 6 from these six values (*i. e.*, first value from first value, etc.), one obtains the dummy observations,

—1.424, —1.205, —1.730, —3.405, —1.242, —2.826

Thus, for the case considered,

$\mathbf{1} = -3.405$ $\mathbf{2} = -2.826$ $\mathbf{3} = -1.730$
$\mathbf{4} = -1.424$ $\mathbf{5} = -1.242$ $\mathbf{6} = -1.205$

To investigate whether $\mu - \nu < -1$ at the 1.6% significance level (this is equivalent to testing $\nu - \mu > 1$), let us use the test

*Accept* $\mu - \nu < -1$ *if* $(\mathbf{6}) < -1$.

Since $-1.205 < -1$, the value of $\mu - \nu$ is significantly less than —1

(or equivalently, $\nu - \mu$ significantly $> 1$) at the 1.6% significance level.

Using

$$[.63(\mathbf{1}) + .37(\mathbf{2}), \quad .63(\mathbf{6}) + .37(\mathbf{5})]$$

for $n = 6$, it is seen that

$$(-3.191, \quad -1.219)$$

is a symmetrical confidence interval for $\mu - \nu$ with a confidence coefficient of approximately 95%. Thus

$$(1.219, \quad 3.191)$$

is a symmetrical confidence interval for $\nu - \mu$ with confidence coefficient nearly equal to 95%; *i. e.*, the probability that the true value of $\nu - \mu$ lies between 1.219 and 3.191 is approximately .95.

5. *Power efficiency investigation.* Let the $m$ observations be a sample from the normal population with mean $\mu$ and variance $\sigma_1^2$, denoted by $N(\mu, \sigma_1^2)$, while the $n$ observations are an independent sample from $N(\nu, \sigma_2^2)$. The only situations considered will be those for which $m = rn$. Cases for which $m = rn + s$, $(s \neq 0)$, are much more difficult to analyze. Moreover, the efficiency for these cases can be estimated from the results for $s = 0$ by interpolation. The only generalized tests of $\mu - \nu < \phi_0$ investigated will be those based on the tests of [5, Tables 1 and 2] for which approximate power efficiencies for normality are listed.

First, let us consider case (a). The $n$ dummy observations on which the solution is based (see Section 3) represent a sample from a normal population. Comparison with [1] shows that the $t$-test formed from this sample is equivalent to Scheffé's "best" $t$-test solution for the Behrens-Fisher problem. Thus the power efficiency of a generalized test for case (a) equals the power efficiency (for normality) of the test from [3] or [5] on which the generalized test is based. For example, if $n = 6$ and the generalized test used is

$$\textit{Accept } \mu - \nu < \phi_0 \textit{ if } \max[\mathbf{5}, (\mathbf{4} + \mathbf{6})/2] < \phi_0,$$

the power efficiency of this generalized test for case (a) is the efficiency listed for the test

$$\textit{Accept } \phi < \phi_0 \textit{ if } \max[x_5, (x_4 + x_6)/2] < \phi_0$$

in [5, Table 1], namely 97%. Thus, the generalized tests are almost as powerful as Scheffé's "best" $t$-test solution for small values of $n$. It

is to be observed that the efficiencies obtained for case (a) are independent of $r$. Table 2 contains a list of the resulting power efficiencies for the tests investigated.

TABLE 2

*Some efficiencies for cases (a) and (b)*

| $n$ | TEST: Accept $\mu - \nu < \phi_0$ if given statistic $< \phi_0$ | SIGNIFICANCE LEVEL (approximate) | APPROXIMATE EFFICIENCY: Case (a) % | Case (b) $r=1$ % | Case (b) $r=\infty$ % |
|---|---|---|---|---|---|
| 4 | **(4)** | .06 | 95 | 77 | 55 |
| | 1.055(**4**)—.055(**1**) | .05 | 96 | 76 | 50 |
| 5 | .63(**5**)+.37(**4**) | .05 | 99 | 83 | 65 |
| | (**5**) | .03 | 96 | 76 | 52 |
| 6 | max[**5**,(**4**+**6**)/2] | .05 | 97 | 84 | 69 |
| | 1.06(**6**)—.06(**1**) | .01 | 98 | 73 | 44 |
| 7 | max[**5**,(**4**+**7**)/2] | .05 | 95 | 84 | 74 |
| | .785(**7**)+.215(**6**) | .01 | 97 | 76 | 52 |
| 8 | max[**6**,(**4**+**8**)/2] | .04 | 94 | 84 | 74 |
| | max[**7**,{.5(**8**)+.28(**6**)+.22(**7**)}] | .01 | 98 | 79 | 59 |
| 9 | max[**6**,(**4**+**9**)/2] | .05 | 91 | 82 | 74 |
| | max[**8**,(**5**+**9**)/2] | .01 | 96 | 77 | 61 |
| 10 | max[**6**,(**4**+**10**)/2] | .06 | 88 | 80 | 73 |
| | max[**8**,(**6**+**10**)/2] | .01 | 96 | 82 | 67 |
| 11 | max[**7**,(**5**+**11**)/2] | .03 | 89 | 80 | 70 |
| | max[**9**,(**7**+**11**)/2] | .005 | 97 | 81 | 64 |
| 12 | max[**9**,(**6**+**12**)/2] | .01 | 94 | 82 | 69 |
| 13 | max[**10**,(**7**+**13**)/2] | .005 | 94 | 81 | 67 |
| 14 | max[**10**,(**6**+**14**)/2] | .01 | 90 | 80 | 70 |
| 15 | max[**11**,(**7**+**15**)/2] | .005 | 92 | 80 | 68 |

For case (b), it is required to find the value $u$ such that the most powerful one-sided test of $\mu - \nu < \phi_0$ ($\sigma_1^2/\sigma_2^2$ known) based on $ur$ sample values from $N(\mu, \sigma_1^2)$ and $u$ sample values from $N(\nu, \sigma_2^2)$ will have approximately the same power function as the corresponding one-sided generalized test based on $nr$ and $n$ sample values, respectively. The power efficiency of the generalized test is then $100u/n\%$ for case (b).

If the ratio of variances $\sigma_1^2/\sigma_2^2$ is known, the most powerful significance test for $\mu - \nu < \phi_0$ is a $t$-test with $m + n - 2$ degrees of freedom

(see [1, p. 43]). The power function values $\epsilon$ for this $t$-test are approximately given by the relation

$$K_\epsilon = K_\alpha - [\phi_0 - (\mu - \nu)]\{1 - K_\alpha^2/2[(r+1)u - 2]\}^{\frac{1}{2}}/\sqrt{\sigma_1^2/ru + \sigma_2^2/u},$$

where $K_\xi$ is the value of the standardized normal deviate exceeded with probability $\xi$ and $\alpha$ is the significance level. This approximation is obtained by modification of the normal approximation to the power function of a one-sided $t$-test given in [6].

Due to the transitivity property of power function equivalence (see [3]), a one-sided generalized test of $\mu - \nu < \phi_0$ based on a test of [3] or [5] with an efficiency of $100E\%$ (for normality) has approximately the same power function as a one-sided $t$-test of $\mu - \nu < \phi_0$ (same significance level) based on $nE$ sample values from a normal population with mean $\mu - \nu$ and variance $\sigma_1^2/r + \sigma_2^2$. Again using the power function approximation for one-sided $t$-tests, the power function values $\epsilon'$ of the generalized test are approximately determined by

$$K_{\epsilon'} = K_\alpha - [\phi_0 - (\mu - \nu)][1 - K_\alpha^2/2(nE - 1)]^{\frac{1}{2}}/\sqrt{\sigma_1^2/rnE + \sigma_2^2/nE},$$

where $\alpha$ is the significance level. Equating $K_\epsilon$ and $K_{\epsilon'}$, the value of $u$ is determined by the equation

$$u\{1 - K_\alpha^2/2[(r+1)u - 2]\} = nE[1 - K_\alpha^2/2(nE - 1)],$$

where all quantities except $u$ have known values. Letting

$$A = A(n, \alpha, E) = nE[1 - K_\alpha^2/2(nE - 1)],$$

and solving this equation for the appropriate root, one obtains

$$u = \frac{1}{2(r+1)}\{2 + (r+1)A + K_\alpha^2/2 + \sqrt{[2 + (r+1)A + K_\alpha^2/2]^2 - 8(r+1)A}\}.$$

Thus for case (b) the efficiency of the one-sided generalized test equals

$$\frac{50}{(r+1)m}\{2 + (r+1)A + K_\alpha^2/2 + \sqrt{[2 + (r+1)A + K_\alpha^2/2]^2 - 8(r+1)A}\}\%.$$

If $r$ is large, this efficiency is approximately

$$100E[1 - K_\alpha^2/2(nE - 1)]\%.$$

The method used to obtain the power efficiency for case (b) is

similar to that used in [7] once the $t$-test which is power function equivalent to the given generalized test is determined.

Table 2 contains power efficiencies of some generalized tests based on [5, Tables 1 and 2] for $r = 1$, $r = \infty$. It is seen that for case (b) the efficiency of the generalized solution is not very great even for $r = 1$.

## REFERENCES


[1] HENRY SCHEFFÉ. On solutions of the Behrens-Fisher problem based on the $t$-distribution. *Annals of Math. Stat.*, Vol. 14 (1943), pp. 35-44.

[2] B. L. WELCH. Generalization of "Student's" problem when several different population variances are involved. *Biometrika*, Vol. 34 (1947), pp. 28-35.

[3] JOHN E. WALSH. Some significance tests for the median which are valid under very general conditions. *Annals of Math. Stat.*, Vol. 20 (1949), pp. 64-81.

[4] JOHN E. WALSH. On the "information" lost by using a $t$-test when the population variance is known. *J. Am. Stat. Assoc.*, Vol. 44 (1949), pp. 122-25.

[5] JOHN E. WALSH. Applications of some significance tests for the median which are valid under very general conditions. *J. Am. Stat. Assoc.*, Vol. 44 (1949), pp. 342-55.

[6] N. L. JOHNSON and B. L. WELCH. Applications of the non-central $t$-distribution. *Biometrika*, Vol. 31 (1940), p. 376.

[7] JOHN E. WALSH. On the power function of the "best" $t$-test solution of the Behrens-Fisher problem. *Annals of Math. Stat.*, Vol. 20 (1949), pp. 616-18.

# AN APPLICATION OF GENE FREQUENCY ANALYSIS TO THE INTERPRETATION OF DATA FROM TWINS

BY DAVID C. RIFE
*Ohio State University*

MONOZYGOTIC twins provide unique material for the study of the relative roles of heredity and environment, and have been utilized extensively for this purpose by many investigators. Numerous traits of a quantitative nature, and for which present techniques of gene frequency analysis are not feasible, have been compared in twins.

The classical method of twin research includes a comparison of the degree of concordance for the trait in question among monozygotic as contrasted with dizygotic twins. If similarities in the environments of the two types of twins are comparable, any greater concordance in monozygotic twins is usually attributed to heredity. Where comparisons are made with respect to traits of a quantitative nature, correlation coefficients may be substituted for percentages of concordance (Newman, Freeman, and Holzinger, 1937). By similar procedures the part played by certain types of environment in bringing about variation may be evaluated by comparisons of monozygotic twins reared together with those reared apart. Actually, monozygotic twins are probably better laboratory material for the student of environment than for the geneticist. We cannot determine modes of inheritance from monozygotic twins alone, although dizygotic twins may be of some value in this respect (Rife, 1938).

Intrapair variations in handedness in both monozygotic and dizygotic twins have long attracted the interest of students of twins. Such intrapair variations in monozygotic twins cannot be attributed to intrapair genotypic differences. Dahlberg (1926), Von Verschner (1927), and Wilson and Jone (1932) assumed that they could be due to circum-

stances *in utero* peculiar to twins. Newman (1937) proposed an asymmetry mechanism to account for mirror-imaging in monozygotic twins, and Lauterbach (1925) proposed that all left-handers were originally members of monozygotic twin pairs. No two investigators used identical criteria for handedness, and the percentages of left-handedness encountered in the two types of twins varied somewhat. Most investigators agree in finding approximately the same percentages of intrapair variations in handedness in the two types of twins. Moreover, the distributions of handedness among both types of twins is very close to what one should expect in a random distribution. If one should draw conclusions solely on the basis of the distributions and relative degrees of concordance in the two types of twins, he might conclude that handedness is entirely a matter of environment.

Evidence from other sources, however, strongly indicates that handedness is at least partially genetic in origin. Various investigators, (Bettman, 1932; Newman, 1934; Cummins, 1940; Rife, 1943b) agree in finding certain correlations between dermatoglyphics and functional handedness. In general, there is a slight reduction in the degree of bimanual asymmetry of left-handers compared with that of right-handers. There is also evidence of linkage between genes responsible for handedness and for interdigital palmar patterns (Rife, 1943a). Both of these relationships between handedness and dermatoglyphics are of a subtle nature and are apparent only when large groups are compared. Yet the observation that *any* relationships exist between handedness and dermatoglyphics is highly suggestive of the operation of genetic factors, in view of the fact that dermatoglyphic configurations are fully established by the end of five months of fetal development.

There is a correlation between the handedness of parents and offspring. When both parents are right-handed, about six per cent of the children are left-handed; when one parent is right-handed and the other left-handed, about seventeen per cent of the children are left-handed, and when both parents are left-handed, about fifty per cent of the children are left-handed. Moreover, the incidence of left-handedness varies from one population to another, presumably due to random drift in gene frequencies (Rife, 1948). There seems to be little doubt that genetic factors play a role in the determination of handedness. We shall now proceed to postulate a hypothesis to account for the incidence of the various combinations of handedness within twins, and shall then demonstrate by gene frequency analysis how certain types of interaction

between heredity and environment may result in almost as much discordance in monozygotic as in dizygotic twins.

## TWIN DATA

We have collected data over a period of years on the handedness of 343 pairs of monozygotic and 211 pairs of dizygotic twins, and the immediate relatives of both groups. These data were taken from twins living in Columbus, Ohio, and at twin conventions in Waterville, Maine; Fort Wayne, Indiana; Waco, Texas; Grand Rapids, Michigan, and Columbus, Ohio. The twins were not selected for handedness, and data for many other traits, including blood groups, types, dermatoglyphics, and photographs were collected from each pair. There is a definite tendency for monozygotic twins to attend such conventions in greater numbers than dizygotic twins, probably because of the greater chances of monozygotic twins attracting attention and winning prizes. We know of no reason, however, to suspect any selection with respect to handedness. This is confirmed by the observation that ratios of handedness types among twins attending conventions approximate those we obtained by canvassing the city schools of Columbus.

Our criteria of handedness included the preferred hand for each of the following acts: throwing, writing, sewing, bowling, shooting marbles, driving a nail, sewing, using a tennis racket, whittling a stick and striking a match. Those using the left hand for one or more of these operations were classed as left-handed, on the assumption that ordinary environment favors right-handers. Approximately two-thirds of those classed as left-handers performed all ten operations with the left hand.

In our early studies of twins, it was observed that pairs showing differences in handedness appeared to have more left-handers in their immediate families (parents and sibs) than did those pairs consisting of two right-handers. The results shown in Tables 1 and 2 indicate that these observations were correct. The dermatoglyphics of pairs showing differences in handedness, as well as those of pairs of left-handers, showed the characteristic trends associated with left-handedness, when compared with those of right-handed pairs. This has been reported in detail elsewhere (Rife, 1943a). These findings indicated that monozygotic twins showing intrapair differences in handedness are genotypically near a threshold, and may be easily shifted by circumstances *in utero* in such a way that they will differ in handedness. An asym-

TABLE 1

*The frequencies of left-handers among immediate relatives of monozygotic twins, classed according to handedness*

In this and the following tables, L signifies left-handed; R-R, twins in which both members are right-handed; R-L, twins showing intrapair differences in handedness; L-L, twins in which both members are left-handed.

| | MONOZYGOTIC TWINS | | | | | | |
|---|---|---|---|---|---|---|---|
| | R-R | | R-L | | L-L | | |
| | Obs. | Cal. | Obs. | Cal. | Obs. | Cal. | Total |
| With L relatives | 44 | 59.40 | 32 | 17.27 | 2 | 1.33 | 78 |
| Without L relatives | 217 | 201.6 | 44 | 58.73 | 4 | 4.67 | 265 |
| Total | 261 | 261 | 76 | 76 | 6 | 6 | 343 |

$df = 2, \quad \chi^2 = 21.74, \quad P < .01$

TABLE 2

*The frequencies of left-handers among the immediate relatives of dizygotic twins, classed according to handedness*

| | DIZYGOTIC TWINS | | | | | | |
|---|---|---|---|---|---|---|---|
| | R-R | | R-L | | L-L | | |
| | Obs. | Cal. | Obs. | Cal. | Obs. | Cal. | Total |
| With L relatives | 26 | 37.31 | 21 | 10.27 | 1 | 0.45 | 48 |
| Without L relatives | 138 | 126.69 | 24 | 34.73 | 1 | 1.55 | 163 |
| Total | 164 | 164 | 45 | 45 | 2 | 2 | 211 |

$df = 2, \quad \chi^2 = 17.28, \quad P < .01$

metry mechanism would be functional within heterozygous pairs. Among dizygotic pairs the heterozygous members may be either right- or left-handed. Pairs genotypically right-handed or left-handed are phenotypically alike.

## GENE FREQUENCY ANALYSIS

The simplest assumption with respect to this hypothesis is that a single pair of genes is responsible for the genetic basis of variations in handedness. This can be readily tested by gene frequency analysis.

Thus, there are the three following genotypes:

RR—right-handed,

Rr—ambidextrous—such monozygotic twins show intrapair differences, and

rr—left-handed.

Let $p$ = frequency gene $R$,
$q$ = frequency gene $r$.

Table 3 shows the percentages of the handedness types among monozygotic twins. Note that $p + q$ does not deviate significantly from 1.

TABLE 3

*Gene frequency analysis of handedness in monozygotic twins*

| HANDEDNESS | NUMBERS | PERCENTAGES | GENOTYPES |
|---|---|---|---|
| R-R | 261 | 76.09 | RR ($p^2$) |
| | | | $p = \sqrt{.7609} = .887$ |
| R-L | 76 | 22.16 | Rr ($2pq$) |
| | | | $q = \sqrt{.0175} = .132$ |
| L-L | 6 | 1.75 | rr ($q^2$) |
| | | | $p + q = 1.019 \pm 0.018$ |
| Total | 343 | | |

If this be the true genetic basis, we can obtain the gene frequencies in a population simply by counting right- and left-handers among the monozygotic twins in a population. Similarly, if we assume that in dizygotic pairs where both are heterozygous, one will be right-handed and the other left-handed, and that where only one member is heterozygous,

half will be right-, and half left-handed, we could obtain gene frequencies by direct count of the twin population.

$$\begin{aligned}\text{Thus, left-handers} &= pq + q^2 \\ &= (1-q)q + q^2 \\ &= q.\end{aligned}$$

This assumption may be tested as shown in Table 4.

TABLE 4

*Gene frequency analysis of handedness in dizygotic twins*

| MATING | FREQUENCIES | HANDEDNESS OF TWINS R-R | R-L | L-L |
|---|---|---|---|---|
| RR × RR | $p^4$ | All | | |
| RR × Rr | $4p^3q$ | 1/2 | 1/2 | |
| Rr × Rr | $4p^2q^2$ | 3/16 | 5/8 | 3/16 |
| RR × rr | $2p^2q^2$ | | All | |
| Rr × rr | $4pq^3$ | | 1/2 | 1/2 |
| rr × rr | $q^4$ | | | All |

$$\begin{aligned}\text{Total R-R pairs} &= p^4 + 2p^3q + 3p^2q^2/4 \\ &= p^2(p^2 + 2pq + 3q^2/4) \\ &= p^2(p+q)^2 - p^2(q^2/4) \\ &= p^2 - p^2q^2/4\end{aligned}$$

$$\begin{aligned}\text{Total R-L pairs} &= 2p^3q + 2.5p^2q^2 + 2p^2q^2 + 2pq^3 \\ &= 2pq(p^2 + 2.25pq + q^2) \\ &= 2pq(p+q)^2 - 2pq(pq/4) \\ &= 2pq + p^2q^2/2\end{aligned}$$

$$\begin{aligned}\text{Total L-L pairs} &= q^4 + 2pq^3 + 3p^2q^2/4 \\ &= q^2(q^2 + 2pq + 3p^2/4) \\ &= q^2(p+q)^2 - q^2(p^2/4) \\ &= q^2 - 2p^2q^2/4\end{aligned}$$

Table 5 shows the actual and expected proportions of the three possible combinations of handedness within dizygotic twins. Note the close agreement between observed and calculated. Of greater significance is the fact that if such a mechanism is indeed responsible for certain

intrapair variations in twins, we should never expect much more disconcordance in dizygotic than in monozygotic twins. The maximum difference, when $p = .5$ and $q = .5$, would only be 3.1125 per cent.

The handedness of twins appears to be more definitely fixed at birth than is the handedness of the single born. Although varying degrees of left-handedness or lack of preference are found, we have yet to encounter a pair of twins, both of whom show no preference as to right or left hand for all of the items included among the tests. Most investigators of handedness in twins have observed higher frequencies of

TABLE 5

*The observed and calculated frequencies of handedness types in dizygotic twins*

The calculated frequencies are based upon formulae derived from Table 4. ($p = .88$, $q = .12$).

| HANDEDNESS | OBSERVED | CALCULATED |
|---|---|---|
| R-R | 164 | 162.81 |
| R-L | 45 | 45.77 |
| L-L | 2 | 2.42 |
| Total | 211 | 211 |

$\chi^2 = 0.115$, $df = 2$, $P$ is between .95 and .70

left-handedness among twins than among the single born. Presumably, this is because half of heterozygous twins are phenotypically left-handed. Most of the heterozygous single born become phenotypic right-handers, as the result of post-natal circumstances. In other words, the gene responsible for left-handedness has higher penetrance among twins than among the single born. It is for this reason we cannot obtain the gene frequencies by determining the ratios of right-handers to left-handers so readily among the single born as among twins.

There is an alternate possibility concerning the phenotypic expression of the interaction of heredity and environment in twins. If we assume that it is a matter of pure chance in twins as to the phenotype of heterozygous persons, we should then expect only half of heterozygous monozygotic twins to show intrapair variations, or $pq$ instead of $2pq$. Obviously, this deviates too much from the observed in handedness to

account for the phenomenon. The expected proportion in dizygotic twins showing intrapair variations could be readily determined by the formulae derived in Table 6. Here, too, the observed proportions of the various combinations of handedness deviate significantly from the expected (Table 7). Note that the percentage of dizygotic twins

**TABLE 6**

*Alternate method of gene frequency analysis in dizygotic twins*

| | R-R | R-L | L-L |
|---|---|---|---|
| $p^4$ | All | | |
| $4p^3q$ | 9/16 | 6/16 | 1/16 |
| $6p^2q^2$ | 1/4 | 1/2 | 1/4 |
| $4pq^3$ | 1/16 | 6/16 | 9/16 |
| $q^4$ | | | All |

Total R-R pairs $= p^4 + 9p^3q/4 + 3p^2q^2/2 + pq^3/4$
$= p(p^3 + 9p^2q/4 + 3pq^2/2 + q^3/4)$
$= p(p^2 + 2pq + q^2)(p + q/4)$
$= p^2 + pq/4$

Total R-L pairs $= 3p^3q/2 + 3p^2q^2 + 3pq^3/2$
$= pq(3p^2/2 + 3pq + 3q^2/2)$
$= pq(3p + 3q)(.5p + .5q)$
$= 1.5pq$

Total L-L pairs $= pq^3/4 + 3p^2q^2/2 + 9pq^3/4 + q^4$
$= p(q^3 + 3pq^3/2 + 9p^2q/4 + p^3)$
$= q(p^2 + 2pq + q^2)(q + p/4)$
$= q^2 + pq/4$

**TABLE 7**

*The observed and calculated frequencies of handedness types in dizygotic twins*

The calculated frequencies are based upon formulae derived in Table 6.
($p = .88$, $q = .12$)

| HANDEDNESS | OBSERVED | CALCULATED |
|---|---|---|
| R-R | 164 | 168.97 |
| R-L | 45 | 33.42 |
| L-L | 2 | 8.61 |
| Total | 211 | 211 |

$\chi^2 = 9.23$, $df = 2$, $P = .01$

showing intrapair differences is theoretically 1½ times greater than in monozygotic twins, according to this hypothesis. While evidently not the basis for intrapair differences in handedness, such a type of interaction may be operative in the expression of many traits.

## DISCUSSION

Monozygotic twins manifest varying degrees of intrapair similarity. Remarkable instances of concordance in regard to the time of onset and degree of severity of various diseases are on record. Yet there are also instances in which one member of the pair has had one of these same diseases, while the other member of the pair was unaffected. Some pairs of unquestionably monozygotic twins show greater intrapair similarities in mental traits, general appearance and personality than do others. Such differences in degrees of intrapair similarity are usually attributed to various environmental circumstances, which are doubtless responsible for many of them. The possibility that *interpair* genotypic differences may also account for variations in the degree of *intrapair* similarity is frequently overlooked. That is to say, the extent to which a trait may be conditioned by environment is dependent on genotype. Twins heterozygous for handedness, for example, are more affected by prenatal circumstances than are those who are homozygous. It is conceivable that part of the variations in the expression of many such intangible traits as vigor, resistance to disease and variations in mental traits may fall into this category.

We have proposed two types of interaction of heredity and environment which may result in intrapair variations in monozygotic twins. The first type results in all heterozygous pairs showing intrapair variations in monozygotic twins, and also in dizygotic pairs in which both members are heterozygous. Monozygotic twins show approximately the same degree of discordance as dizygotic twins where this type of interaction is concerned. Data on handedness in twins conform very well to this explanation.

The second type of interaction results in only one-half of heterozygous monozygotic and in three-fourths of heterozygous dizygotic pairs showing intrapair variations. Degrees of concordance and correlation in twins are conditioned by the frequencies of the genes concerned and their modes of expression, as well as by the relative roles of heredity and environment. It may be misleading, therefore, to estimate the role

of heredity solely on the relative proportions of concordance or correlation in the two types of twins.

SUMMARY

Methods of gene frequency analysis for two types of interaction of heredity and environment in twins are given. Where such interactions are involved, the gene frequencies of a population may be obtained by simply determining the ratio of the alternate phenotypes among the twins, regardless of zygosity.

## REFERENCES

BETTMAN, S. 1932. Die Papillarleistenzeichnung der Handflache in ihrer Beziehhung zur Handigkeit. *Zeitschr. f. Anat. u. Entwick.*, 98: 149-174.

CUMMINS, HAROLD. 1940. Finger prints correlated with handedness. *Amer. Jour. Phys. Anthr.*, 26: 151-166.

DAHLBERG, G. 1926. Twin births and twins from a hereditary point of view. Tidens forlag. (*Stockholm*).

LAUTERBACH, C. E. 1925. Studies in twin resemblance. *Genetics*, 10: 525-568.

NEWMAN, H. H., F. N. FREEMAN and K. J. HOLZINGER. 1937. Twins. University of Chicago Press (*Chicago*).

RIFE, D. C. 1938. Simple modes of inheritance and the study of twins. *Ohio Jour. Sc.*, 37: 281-293.

———. 1943a. Genetic interrelationships of dermatoglyphics and functional handedness. *Genetics*, 28: 41-48.

———. 1943b. Handedness and dermatoglyphics in twins. *Human Biology*, 15: 45-54.

———. 1948. Genetic variability within a student population. *Am. Jour. Phys. Anthr.*, 6: 47-62.

VERSCHUER, O. 1937. Der Vererbungsbiologische Zwillings forschung Ergeb. Innern. Med. u. Kinderheit., 31.

WILSON, P. T. and H. E. JONES. 1932. Left-handedness in twins. *Genetics*, 17: 560-571.

# SEXUAL GROWTH OF NEGRO AND WHITE BOYS

BY GLENN V. RAMSEY

*University of Texas*

THIS study presents data regarding the age of the first ejaculation, age of first appearance of pigmented pubic hair, and age of first recognition of voice change as reported by 37 Negro boys and 286 white boys. All data presented in this study were obtained by personal interviews with the boys. The Negro boys ranged in age from 11 to 16 years; the white boys from 10 to 20, 85 per cent from 12 to 16. A more extensive report on the sexual growth and behavior of the white group has been published elsewhere (2).

The 286 white boys lived in a Middle Western city of over 100,000. Most of the boys were enrolled in the seventh, eighth, or ninth grades of the public or parochial schools of the city. The population studied was fairly representative of the various economic levels of the city. The data on this group were collected during the years 1940 and 1941.

The 37 Negro boys included in this study constituted the *entire* male enrollment in the sixth, seventh, and eighth grades of an exclusively Negro grade school. This school was located in an Eastern residential town with a population of approximately 8,000. The principal occupational pursuits of the Negro families in this town were the domestic and business services. The socio-economic level of the group could be described as middle class for the general Negro population in the Northeastern part of the United States. All the data regarding these Negro boys were obtained during 1948.

## TREATMENT OF DATA

Age was obtained to the closest year and month that the individual could report it. These age reports were then rounded off to the nearest half-year interval. Percentages were calculated for the number of Negro and white males at each age level who had experienced first

ejaculation, and so on. Then medians for each growth phenomena being investigated were calculated. Comparisons are made between data for the Negro and white groups.

## RESULTS

Year intervals for age of first ejaculation were calculated. Then the percentage of boys at each age level who had experienced first ejaculation was calculated. The results are presented in Fig. 1.

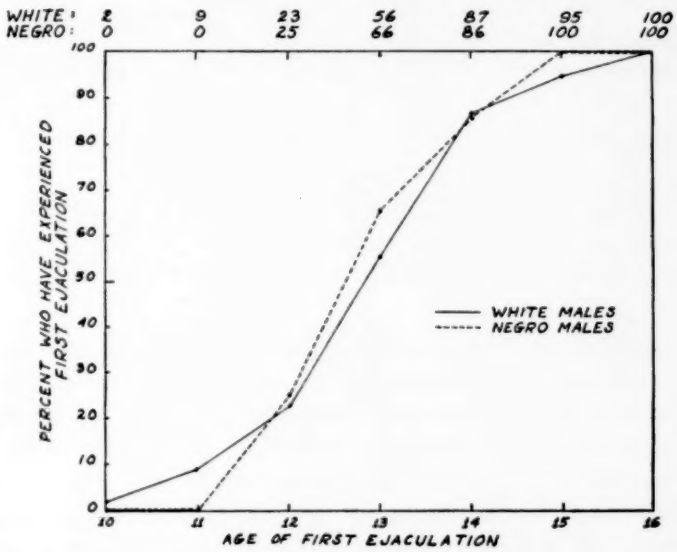


FIG. 1. PERCENTAGE OF 37 NEGRO AND 286 WHITE MALES AT AGE LEVELS FROM 10 THROUGH 16 WHO HAVE EXPERIENCED FIRST EJACULATION

The data for both groups when presented graphically produce the ogive "growth curve." The curve for the white population has a more gradual onset and termination than the curve for the Negro population. This difference possibly results from the larger population which constituted the white sample.

The median age of first ejaculation, first appearance of pubic hair and age of first recognition of voice-change was also calculated for both the Negro and white groups. The median age was calculated from age taken at the nearest half-year interval. These median scores are

shown in Table 1. Comparable data on these three variables taken from the Kinsey report (1) are also included in Table 1.

TABLE 1

*Median age for first ejaculation, first voice-change, and first appearance of pubic hair*

| | RAMSEY DATA | | KINSEY DATA |
|---|---|---|---|
| | Negro N = 37 | White N = 286 | White N = 2500 to 3600 |
| Median age of first ejaculation | 13.8 | 13.8 | 13.77 |
| Median age of first appearance of pubic hair | 13.3 | 13.6 | 13.43 |
| Median age for first recognition of voice-change | 13.7 | 13.4 | 14.44 |

The evidence in Table 1 indicates that for the groups studied there is very little difference between the Negro and white males in regard to the three median measures of sexual growth. The data presented in this study are in close agreement with those provided by Kinsey *et al.* (1), especially on median age of first ejaculation and first appearance of pubic hair. The data presented in this study on voice-change for both Negro and white males show a slightly lower age for onset than that reported by Kinsey. Since voice-change is not a sudden and definite transition point but extends over a period of years, some variability can be expected in reports by various age groups on its first appearance.

The three indices of sexual growth employed in this study reveal that there is very little, if any, difference between the Negro and white groups. More extensive studies than this one will be needed before generalizations can be validly made. The possible influence of other variables, such as climate and nutrition, upon indices of sexual growth might also be worthy of investigation.

SUMMARY

The median age reported for first ejaculation by 37 Negroes was 13.8 years and for 286 whites was also 13.8. The median age for first appearance of pubic hair for the Negro group was 13.3, for the white group, 13.6. The median age for first recognition of voice-change was 13.7 for the Negro males and 13.4 for the white males.

## BIBLIOGRAPHY

1. KINSEY, A. C. *et al.* Sexual Behavior in the Human Male. *Philadelphia*: Saunders, 1948.
2. RAMSEY, G. V. The sexual development of boys. *Amer. J. Psych.*, 1943, 56: 217-234.